

THIS WEEK

EDITORIALS

CLIMATE Careful conditions should allow geoengineering work **p.8**

WORLD VIEW The dangers of partisanship in science policy **p.9**



DECLINE Fussy bees could hunger after long-gone flowers **p.10**

The uncertain dash for gas

The United States and other countries have made huge investments in fracking, but forecasts of production may be vastly overestimated.

A decade ago, fracking was a misprint. Now, hydraulic fracturing is heralded as the future of energy supplies. Leading the way is the United States, which is tapping its rich shale deposits to produce more natural gas than at any other time in its history. The International Energy Agency projected in November that global production of shale gas would more than triple between 2012 and 2040, as countries such as China ramp up fracking of their own shale formations.

How much gas and oil is down there? Predictions about future shortages and abundances of various fuels are tricky, because although the geological presence of resources can be surveyed to some degree, how much would be profitable to extract is a moving target. Academic journals are filled with earnest projections about future energy dynamics, which usually turn out to be wildly inaccurate. Even worse, governments and companies wager billions of dollars on dubious bets. This matters because investment begets further investment. As the pipework and pumps go in, momentum builds. This is what economists call technology lock-in.

This week, *Nature* presents previously unpublished data suggesting that the lock-in of technology into shale-gas production may be a riskier bet than previously realized, at least in the United States. *Nature* has obtained detailed US Energy Information Administration (EIA) forecasts of production from the nation's biggest shale-gas production sites. These forecasts matter because they feed into decisions on US energy policy made at the highest levels. Crucially, they are much higher than the best independent academic estimates.

The full story is contained in a News Feature on page 28. The conclusion is that the US government and much of the energy industry may be vastly overestimating how much natural gas the United States will produce in the coming decades.

The EIA projects that production will rise by more than 50% over the next quarter of a century, and perhaps beyond, with shale formations supplying much of that increase. But such optimism contrasts with forecasts developed by a team of specialists at the University of Texas, which is analysing the geological conditions using data at much higher resolution than the EIA's. The Texas team projects that gas production from four of the most productive formations will peak in the coming years and then quickly decline. If that pattern holds for other formations that the team has not yet analysed, it could mean much less natural gas in the United States' future.

Like all energy forecasts, the lower projections from the Texas team could turn out to be inaccurate. Technological advances in the next few decades could open up more resources at lower costs, driving US production even higher than the EIA has predicted. But it is also possible that the Texas forecasts are too high, and that gas production will fall off even faster than the team suggests.

The one certainty here is that the United States and other nations have invested relatively little in tracking and assessing their natural resources. The EIA has a total budget of US\$117 million, less than the value of one

day's gas production from the country's shale formations. The agency's budget has increased in recent years, but US natural-gas production has grown much more rapidly. There are now tens of thousands of wells tapping shale formations, and thousands more are started each year.

US federal and state governments have also failed to keep up with environmental concerns associated with fracking of shale formations for gas. For each well, drillers use tens of millions of litres of water, as well as untold amounts of chemical additives — and most states do not require them to report which compounds they use. Environmental groups have charged that fracking fluids can contaminate drinking water near wells, and emissions associated with drilling and production have caused air pollution. There have been few government or academic studies so far, but some

research has found evidence of contamination associated with shale-gas production. Similar environmental problems plague fracking of shale deposits for oil, which has led to a sharp rise in US petroleum production in the past five years. And then there is the bigger picture: the extra greenhouse-gas emissions that come from new sources of fossil fuels.

Fracking has momentum. It will probably continue to grow to become an important part of the energy mix in many parts of the world. But for strategic, economic and environmental reasons, all involved should take a hard look at the numbers. ■

"Nations have invested relatively little in tracking and assessing their natural resources."

Harsh reality

Two reports highlight the plight of postdocs on both sides of the pond aiming for academia.

When hundreds of UK scientists were asked in a recent exercise to describe high-quality scientific research, the most popular word that they suggested was "rigorous". Most were probably referring to its dictionary definition as "extremely thorough and careful". But more than a few must have been aware that the word has some other, equally valid, synonyms: rigidly severe, harsh, tough.

Good science is tough. But is it also harsh and severe? And if so, does it need to be? At what point do the legitimate demands of competitive academic research tip into a demoralizing lack of job security and intolerable pressure? It has been said before, not least in these pages, but two reports published this week on either side of the Atlantic highlight perhaps the most common pinch point: the postdoctoral years.

Although the lament of the postdoc may be a familiar cry, all who care about the current state of science and where it is heading would do well to look at the separate reports, which present a visceral and honest snapshot of opinions from life in the squeezed middle of academia.

The UK report is the work of the Nuffield Council on Bioethics. Earlier this year, it surveyed 970 people involved in research at UK universities and institutions, and held detailed discussions with another 740. Postdocs made up the largest single group, but significant numbers of respondents held more senior posts, right up to heads of department.

The US effort is a write-up of an October seminar held by postdoc researchers in and around Boston, Massachusetts (G. S. McDowell *et al.* *F1000 Res.* <http://doi.org/xg9>; 2014). It is published ahead of a related symposium at the annual meeting of the American Society for Cell Biology in Philadelphia, Pennsylvania, which starts this weekend.

Given a platform to complain, most people will. Both reports grumble about perennial problems that are perceived to run through research. Government funding is insufficient, external focus on journal impact factors stifles creativity, and bureaucracy and distractions mean that everyone has less time to spend on what they really want to do.

These are common legitimate concerns, but how about this: a whopping 58% of scientists in the UK report said that they were aware of colleagues feeling tempted or under pressure to compromise on research integrity or standards. Asked whether they felt this way themselves, just 21% of scientists aged 35 or over said yes; strikingly, that figure shot up to one-third of those aged under 35. In the United States, postdocs consistently called themselves “the lost people” and “the invisible people”. The US report states that “junior scientists are primarily treated as cheap labor rather than as participants in a well-rounded training program”.

It is no longer acceptable for senior scientists to ignore such complaints. Research in 2014 is a brutal business, at least for those who want to pursue academic science as a career. Perhaps the most telling line comes from the UK report: of 100 science PhD graduates, about 30 will go on to postdoc research, but just four will secure permanent academic posts with a significant research component. There are too many scientists chasing too few academic careers.

“Research in 2014 is a brutal business.”

That has been the reality for some time, but the message is yet to penetrate. The US report says that lab heads train scientists “in their own image, that is, for a career in academia, though only a small minority will obtain tenure-track faculty positions”. Postdocs say that an academic career is still presented to them as the default outcome. There is a “complete lack of information on number of postdocs”, notes the US report.

There is a gap between reality and expectations. Ironically for a career that demands dispassionate judgements based on data and evidence, the postdoc experience is too often a leap of faith that leaves bright and talented people disillusioned and directionless.

The solutions are many, but will require time because they demand a change in culture. Postdoc contracts need to be more than an entry-level position for a career that few will follow. Institutions that offer them must be transparent about future prospects and help postdocs to develop transferable skills to ease their transition into the broader job market.

The philosophy can be boiled down to this: it is a good thing, for both the individuals and society at large, that these young people spend some of their most productive years tackling research. And it is a good thing that most take that independence into other occupations. ■

Look ahead

Research into climate engineering must proceed — even if it turns out to be unnecessary.

The irony in discussions about climate engineering is that, while society considers its merits, the process itself is already in full swing. With vast amounts of heat-trapping molecules released each day into the atmosphere, humans are deliberately altering the planet's climate in unpredictable ways. The magnitude of the resulting climate change is worryingly uncertain. Even more uncertain are the physical, social and economic side effects of global warming. There is every reason to believe that, by and large, they will be harmful.

Why, then, is the idea that future generations could use a little science and engineering to deliberately cool the world so controversial? The answer, of course, is that the cure could be worse than the disease.

Adding sulphate into the high atmosphere, for example, is one of a broad range of geoengineering techniques proposed in response to the warming driven by greenhouse gases. If the technique helps to destroy the ozone layer or increases drought risk in vulnerable regions, then there is a strong argument not to do it.

Scientists are not solely responsible for the problem of global warming. And many argue forcefully that they should be wary of simply replacing one evil with another. Even scientists who are directly involved in geoengineering studies often admit that they do not like the prospect of their research becoming a real-world necessity.

There are some aspects of geoengineering on which all can agree. It should not distract from efforts to curb emissions. An effective political agreement to radically reduce greenhouse-gas emissions, such as that being discussed this week at the United Nations climate-change conference in Lima, must take priority over speculative notions to instead tinker with the atmosphere to meet climate goals.

In fact, geoengineering practices that do pose significant further risk to the environment must be prohibited, if necessary by international law. After all, no single nation — let alone any faction of science — can assume the right to deliberately modify the physical set-up of the planet.

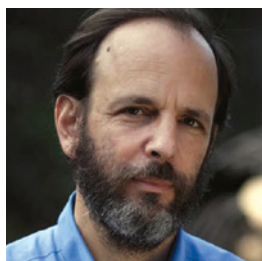
Large-scale and possibly irreversible atmospheric interventions are clearly beyond what is scientifically and ethically justified. But apart from behemoth plans (which nobody is seriously promoting), there are many more limited climate-engineering options that do deserve serious consideration and study. To that end, leading scientific societies are this week discussing a set of guiding principles for responsible field experiments (see page 20). It would greatly enhance the credibility of the field if it could adopt such a scientific code of conduct.

The geoengineering option that seems simplest — scraping carbon dioxide from the air and permanently locking it somewhere secure — is already being intensely investigated. Carbon capture and storage technology is now widely considered to be safe, but technical and financial challenges limit its wide-scale adoption. Because the world's appetite for fossil fuels has not yet peaked, it is as important to encourage and fund research on the carbon capture side of the technology as on the carbon storage aspect. But whether this technology will really help to fight climate change depends on political governance, such as whether it becomes standard in the international energy sector to fit new coal-burning plants with carbon capture equipment.

In its last report, the Intergovernmental Panel on Climate Change (IPCC) left little doubt that some form of geoengineering (or ‘negative emissions’, in IPCC language) will probably be needed to meet the goal of limiting global warming to 2°C. Having delivered its fifth full climate assessment report since 1990, the IPCC is considering adopting a new role in the future. If the group were to switch to more-focused, trimmed-down reports, delivered on demand, a

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunqv

special report on climate engineering might be the perfect place to start. Meanwhile, researchers should work fast to clear the way for more responsible research, even if responsible action means that its results will never be needed. ■



Science should keep out of partisan politics

The Republican urge to cut funding is not necessarily anti-science, and the research community ought not to pick political sides, says Daniel Sarewitz.

Two weeks after US voters installed a Republican majority in the Senate and expanded the Republican majority in the House of Representatives, the American Association for the Advancement of Science (AAAS) made its own political statement: it named a former Democratic member of Congress as its new chief.

Once, these events might have been unrelated. But in today's poisonous partisan atmosphere, the AAAS's choice of Rush Holt, a physicist and political centrist just finishing a 16-year stint in Congress, looks every bit as political as the election itself.

It is standard wisdom among Democrats that Republicans are 'anti-science'. This view will be reinforced when Oklahoma Senator James Inhofe, famously sceptical about climate change, takes over the Committee on Environment and Public Works in January; when House science committee chair Lamar Smith (Texas) renews his assault on social science and the peer-review process; and when research and development spending continues to stagnate under a Republican-controlled Congress.

The AAAS, which bills itself as "the world's largest general scientific society" has positioned itself to counter these developments by anointing a leader who could take up the fight. From this perspective, the choice of Holt might seem inspired. That is certainly what commentators on the Democratic side are saying. Typical is a blog post by Joe Romm of the think tank Center for American Progress in Washington DC, who looks forward to Holt continuing "his blunt defense of both science and climate action given his new high-profile platform".

But is it smart for the AAAS to link itself explicitly to the partisan fray? The generally accepted metric of how well national science is doing is the level of government funding, and by that measure Republicans have, on the whole, supported science as much Democrats have in the past 50 years. The problem today is not that Republicans particularly want to limit science spending. It is that in their obsession with fiscal restraint they are willing to cut all spending, including in areas they have historically favoured, such as military programmes — and basic science.

Continuing Democratic efforts to paint Republicans as congenitally anti-science could backfire. During the height of the Ebola scare, National Institutes of Health (NIH) director Francis Collins told *The Huffington Post* that "if we had not gone through our 10-year slide in research support, we probably would have had a vaccine in time". Republicans interpreted that as blaming them for the impact of the Ebola outbreak, an interpretation backed up by a pre-election television ad sponsored by a pro-Democratic political group that showed pictures of dead Africans and

concluded with the words: "Republican Cuts Kill." Republican politicians responded by suggesting that, with its US\$30-billion budget, the real problem for the NIH was poor priorities. "NIH did come up with the money to pay to give Swedish massages for rabbits," noted Republican Senator Tom Coburn (Oklahoma).

Actually, the idea that science policy can set priorities rather than always push for bigger budgets is sensible, especially in a time of fiscal restraint. But the US science community has never wanted to have this conversation, because it would undermine what has been a fundamental tenet of science policy since the 1950s: that more funding automatically translates into more social benefit. So the conversation has been handed over to politicians.

Smith's attacks on National Science Foundation (NSF) funding for social-science programmes take aim at the belief that the best national science policy is the one that puts the most money into science. For the past year and a half he has been picking out grants that he thought sounded frivolous — a time-honoured political ploy — and asking the NSF to provide the peer reviews that justify the funding decisions. The day after the election, he explained his efforts as a matter of setting priorities: "I support continued funding for worthy social science research projects. But funding for social science should not come at the expense of areas of science ... that are most likely to produce breakthroughs that will save lives, create jobs, and promote economic growth."

One irony here is that if Smith and other Republicans were more familiar with social science, they could use its findings to counter accusations that they are anti-science or less scientifically informed than Democrats. Decades of research have shown that people — including scientists — interpret the real-world implications of science in terms of their cultural background, personal experience and political beliefs. It is no surprise, for example, that conservatives who are distrustful of the ability of government to guide large-scale social change are distrustful of climate scientists and activists who invoke climate science to argue for such change.

The political situation surrounding US science and politics is not clear-cut. The more the AAAS, and so the science community, is seen to line up behind one party, the less claim it will have to special status in informing difficult political and social decisions. Public regard for scientists remains particularly high, and for politicians, particularly low. Blurring the boundaries between these groups is not likely to redound to the benefit of politicians, but to the detriment of scientists. ■

Daniel Sarewitz is co-director of the Consortium for Science, Policy and Outcomes at Arizona State University, and is based in Washington DC. e-mail: daniel.sarewitz@asu.edu

**SCIENCE POLICY
CAN
SET PRIORITIES
RATHER THAN
ALWAYS PUSH FOR
BIGGER
BUDGETS.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/nyom5t

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

QUANTUM COMPUTING

Qubits come close to perfection

Physicists have created quantum units of information that operate with some of the lowest levels of error ever seen.

David Lucas at the University of Oxford, UK, and his colleagues trapped ions of the rare isotope calcium-43 with electric fields and manipulated them with lasers and microwaves, forming 'quantum bits' or qubits. Their set-up was robust enough to outside interference that the qubits kept their fragile quantum states for 50 seconds and errors arose only once in every 1 million operations, 10 times better than previous demonstrations.

Because correcting codes can overcome these very low levels of error, the authors say that their qubits are accurate enough to be used in quantum computing. However, linking up many of these qubits in a scalable system remains a challenge. *Phys. Rev. Lett.* 113, 220501 (2014)

CANCER

Old blood reveals cancer risk factors

DNA sequencing could help to identify people who are at risk of developing blood cancers months or even years before malignancy appears.

To search for early indicators of cancers such as acute myeloid leukaemia, Steven McCarroll at Harvard Medical School in Boston, Massachusetts, and his colleagues sequenced the DNA from stored white blood cells collected from 12,380 people. A team

led by Benjamin Ebert at Brigham and Women's Hospital, also in Boston, looked at similar data for another 17,182 people. Both groups then used medical records to find out which people had subsequently been diagnosed with cancer. Older people were more likely to have mutations in genes implicated in blood cancers. People carrying these mutations were up to 32 times more likely than non-carriers to be diagnosed with a blood cancer months

or years later.

The authors point out that DNA sequencing cannot yet predict blood cancers accurately enough for use in clinical care.

N. Engl. J. Med. <http://doi.org/xdp> (2014); <http://doi.org/xdq> (2014)

MATERIALS

How the silver Koi carp shines

Japanese koi carp are famous for their iridescent shimmer, but Lia Addadi

at the Weizmann Institute of Science in Rehovot, Israel, and her colleagues have discovered why some glimmer more brightly than others.

They used correlated optical and electron microscopy to compare scales and skin of a common variety of koi with those from a shinier type called Gin Rin. In the scales and skin of both types are cells called iridophores, which contain reflective stacks made up of layers of guanine crystals and



BERT PIJS/MINDEN PICTURES/FLPA

CONSERVATION BIOLOGY

Bees lose their favourite flowers

Shrinking populations of bees' preferred plants could help to explain the worldwide decline in many wild bee species.

Jeroen Scheper of Wageningen University in the Netherlands and his colleagues carefully scraped pollen grains off the legs of museum specimens of 57 Dutch bee species to identify which plants the bees had fed on. The bees were collected before 1950, and thus before many wildflower populations began to

decline because of agricultural intensification. The study revealed that bees whose favourite plants have declined the most are now showing the steepest drop in population numbers. Populations of larger bees, which have greater food requirements, also showed big decreases.

The findings support calls to plant specific flowers for these threatened pollinators.

Proc. Natl Acad. Sci. USA <http://doi.org/xdx> (2014)

cytoplasm. The Gin Rin fish had a much greater density of iridophores, each containing more stacks of crystals than in the common type. The Gin Rin crystals were also oriented differently — almost parallel to the surface of their scales instead of at an angle of around 30°, as in the common variety.

J. Am. Chem. Soc. <http://doi.org/xc8> (2014)

GEOLOGY

Europe feels fracking shakes

Hydraulic fracturing during natural-gas extraction caused two earthquakes in the United Kingdom that were felt by residents, by reactivating a geological fault that had lain dormant for many millions of years.

'Fracking' involves injecting a mixture of water, sand and chemical additives into the ground to fracture rock and loosen trapped natural gas. Earthquakes induced by the process in North America have been felt by communities, but this is the first such example in Europe.

A team led by Huw Clarke of fracking company Cuadrilla Resources in Lichfield, UK, reports that fracking in 2011 near Blackpool caused earthquakes on 1 April and 27 May of magnitudes 2.3 and 1.5, respectively. Fluid injected during the fracking

operation (**pictured**) caused movement in a fault that had been inactive for 260 million years.

Geophys. Res. Lett. <http://doi.org/xc9> (2014)

CANCER IMMUNOLOGY

Predicting cancer-therapy success

Tumour-cell genomes may hold clues to who will benefit from treatments that stimulate the immune system against melanoma.

A protein called CTLA-4 on the surface of cancer-fighting T cells can suppress these immune cells' attacks on the tumours. Drugs that block this protein can, in some patients, unleash these cells, but can have toxic side effects. Jedd Wolchok and Timothy Chan of the Memorial Sloan Kettering Cancer Center in New York and their colleagues sequenced genes in tumour cells from 64 patients with melanoma who were treated with CTLA-4 inhibitors.

Patients whose tumours had a higher number of cancer-associated mutations were more likely to benefit from the drugs, and a set of 101 tumour proteins created by these mutations were associated with a strong response to the drugs.

N. Engl. J. Med. <http://doi.org/xdx> (2014)

STEM CELLS

Matched stem cells still rejected

Transplants of stem cells with nuclear DNA identical to that of a patient could still trigger immune responses because of the presence of mismatched mitochondria.

Sonja Schrepfer at Stanford University in California and her colleagues transferred nuclei from a donor mouse's adult cells into mouse embryonic stem cells with genetically matched nuclear DNA but different mitochondrial DNA. When

SOCIAL SELECTION

Popular articles on social media

Unequal fates for maths superstars

A gift for numbers can take a person far in life, according to a report getting plenty of online attention. A survey of 1,004 men and 601 women who were identified as 13-year-old mathematics prodigies in the 1970s found above-average levels of accomplishment in fields including business and academia. "How do things turn out for math prodigies? Pretty well, it seems," tweeted Tyler Cowen, an economist at George Mason University in Fairfax, Virginia. Cowen told *Nature* that he shared the report on Twitter because he suspects that many of his followers have gifted children. "I was delivering good news," he says.

However, success was not evenly distributed. On average, the women in the survey earned about US\$80,000 a year. That is more than double the amount that US women with full-time jobs typically make in a year, but about \$60,000 less than the men in the survey. Zhana Vrangalova, a sexuality researcher at New York University in New York City, summed it up on Twitter by noting that the survey participants had accomplished "a lot more than the average Jo/Jane".

Psychol. Sci. <http://doi.org/xfb> (2014)



Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➔ **NATURE.COM**
For more on popular papers:
go.nature.com/ee3r5w

the cells were transplanted back into the donor, they were attacked by the immune system. Isolated mismatched mitochondria also triggered an immune response.

Cell Stem Cell 16, 1–6 (2015)

CANCER BIOLOGY

Tumours set stage for their spread

An intricate interplay between cancer cells and white blood cells outside a lung tumour can help to drive metastasis, the spread of the disease to other parts of the body.

Nathan Reticker-Flynn and Sangeeta Bhatia at the Massachusetts Institute of Technology in Cambridge focused on a molecule called galectin-3, which they found is expressed on the surface of certain white blood cells during early cancer in mice. Lung tumours secrete signalling molecules that

mobilize these white cells — known to promote metastasis — out of the bone marrow and into the bloodstream.

Metastatic cells from the tumour display a cell-surface sugar that binds to galectin-3. As a result, these cells increasingly interact with the mobilized white blood cells in distant parts of the body, enhancing the cancer cells' ability to colonize and grow into new tumours.

Cancer Discov. <http://doi.org/xdx> (2014)

CORRECTION

The Research Highlight 'Fish tags guide seal predators' (*Nature* 515, 469; 2014), gave the wrong year for the citation; it should have been 2015.

➔ **NATURE.COM**
For the latest research published by Nature visit:
www.nature.com/latestresearch



SEVEN DAYS

The news in brief

POLICY

HIV in Europe

Eastern Europe and Central Asia have seen a surge in new HIV cases in the past decade, the World Health Organization warned in a report released on 27 November. Last year, 136,235 new cases of HIV infection were reported in what the agency calls the European region, with almost 60% of these diagnoses in Russia. The rate of reported new HIV diagnoses in countries in the east of the region, which includes Russia, has shot up by 144%, from 16.9 per 100,000 people in 2004 to 41.2 in 2013.

Smog standards

The US Environmental Protection Agency on 26 November released proposed rules that would tighten standards for smog-producing ozone in the atmosphere. The draft measure would lower the threshold for acceptable levels from 75 parts per billion to between 65 and 70 parts per billion. Agency scientists examined more than 1,000 studies to settle

NATURE CONTENT SHARING

All papers from *Nature* and *Nature* research journals will be shareable in a proprietary screen view that can be annotated, but not copied, printed or downloaded. The policy applies to 49 journals on the nature.com domain, and lets subscribers share links to read-only PDFs of any papers they have access to. About 100 media outlets and blogs will also be able to share a link (which anyone can later repost); readers will be able to save articles using the free desktop software ReadCube. See go.nature.com/7nijn for more.



RAPHAEL ALVES/AFP/GETTY

Amazon logging levels down

Deforestation in the Brazilian Amazon dropped by 18% in 2014, reaching the second-lowest level since monitoring began in 1988, the Brazilian government announced on 26 November. The low-resolution satellite data used for enforcement activities had suggested increased deforestation this year. But high-resolution satellite imaging showed just 4,848 square kilometres of new clearings from

August 2013 to July 2014, nearly 83% below the peak in 2004. Most of the newly cleared land is destined to become cattle pasture. Deforestation hit a record low of 4,571 square kilometres in 2012, but a decade of aggressive law enforcement has sparked a political backlash. The Brazilian Congress weakened the country's forest-protection law in 2012, and deforestation spiked by 29% last year.

on a range projected to better safeguard against asthma and other lung diseases. States would be required to meet the new standards between 2020 and 2037, depending on the severity of their ozone problem.

Beijing smoking ban

The city of Beijing on 28 November adopted a law banning smoking in indoor public spaces. China is considering implementing the same restrictions nationwide, according to the country's state-run news agency, Xinhua. The Beijing law, due to be enforced from June 2015, also bans tobacco advertising. China has tried before to restrict smoking in public spaces such as hotels

and restaurants — but the rules were never rigorously enforced. The country is the world's largest consumer of tobacco, with more than 300 million smokers.

EVENTS

Polio killings

Three polio vaccinators and their driver were killed in Quetta, a city in western Pakistan, on 26 November. It was the latest in a series of fatal attacks on health workers in Pakistan — where the Taliban have banned polio vaccinations — over the past two years. The country has seen a surge in polio cases this year, with 260 reported as of 25 November, compared with 64 at the same time last year.

Probe postponed

Forecasts of bad weather delayed the planned launch on 30 November of Japan's Hayabusa-2 spacecraft. The mission to 1999 JU3, a 900-metre-long asteroid, aims to collect sample material that is relatively unchanged from 4.5 billion years ago, when gas and dust were clumping into particles around the newborn Sun. See go.nature.com/ofm5if for more.

Nobel for sale

James Watson is selling the gold medal that he received as part of his share in the 1962 Nobel Prize in Physiology or Medicine, which he won with Francis Crick and Maurice Wilkins for their work uncovering the double-helix

structure of DNA. Watson's medal is expected to fetch up to US\$3.5 million at an auction on 4 December. Nobel medals have been sold before, including Crick's, which drew \$2.27 million at auction last year. Watson's is the first to be offered by a living recipient.

PEOPLE

Scientist sentenced

A scientist who worked at Sandia National Laboratories in Albuquerque, New Mexico, has been sentenced to a year and one day in federal prison, the US Attorney's Office in New Mexico announced last week. Jianyu Huang was arrested in 2012 on six charges, including the misuse of US government resources and equipment to conduct research for institutions in China. In August, Huang pleaded guilty to unlawfully carrying a US government laptop with him on a trip to China, and to falsely stating beforehand that he would not take the computer.

Iranian minister

Iran's parliament voted on 26 November to accept Iranian president Hassan Rouhani's nominee, Mohammad Farhadi, as science minister. The parliament had previously rejected four candidates, apparently in protest over Rouhani's plans to encourage



greater staff and student freedoms at universities — a sensitive issue because campuses were a hotbed of political protests in 2009 (see go.nature.com/hk7zkh). Farhadi (pictured), an ear, nose and throat researcher at the Iran University of Medical Sciences, Tehran, heads Iran's Red Crescent Society, a medical-aid organization. He was formerly health minister under the reformist president Mohammad Khatami, and science minister in the 1980s under prime minister Mir-Hossein Mousavi.

Plagiarism charge

Deepak Pental, a well-known Indian geneticist and former vice-chancellor of the University of Delhi, has been accused of plagiarism in a case that saw him arrested but quickly released on 25 November. Concerns were raised by P. Pardha Saradhi, an environmental biologist at Delhi, who says that his

work on genetically modified mustard seed was plagiarized by Pental and his postdoc K. V. S. K. Prasad, who did his PhD with Saradhi. Saradhi also says that Pental used material from his laboratory (then at the National Islamic University in New Delhi) without permission. Pental told *Nature* that “the charges are false and frivolous and will not stand the scrutiny of law”. The High Court of Delhi is looking into the case.

FUNDING

Quantum industry

The United Kingdom has picked four research hubs — based at the universities of Oxford, York, Glasgow and Birmingham — to share £120 million (US\$188 million) to develop technologies based on quantum mechanics. Seventeen universities and 132 companies will work across the hubs, which will focus on developing quantum computers; quantum-based imaging devices such as cameras; gravity sensors and atomic clocks; and encryption. See go.nature.com/21puqw for more.

BUSINESS

Renewables shift

One of Europe's biggest energy providers, E.ON, is restructuring to focus

COMING UP

9–11 DECEMBER

The Regional Aquaculture Conference takes place in Bari, Italy, coordinated by groups including the European Commission and the Food and Agriculture Organization of the United Nations. It focuses on sustainable aquaculture in the Mediterranean Sea and the Black Sea.

go.nature.com/lpo1nx

on renewables, electricity distribution networks and energy-efficiency services. The Germany-based company will offload its conventional power generation (coal, gas and nuclear fuel) and energy-trading business into a separate, independent company, it announced on 30 November. The idea is to “propel the transformation toward a clean, sustainable energy supply”, Johannes Teyssen, E.ON's chief executive, told journalists.

Pharma hacked

Dozens of health-care and pharmaceutical firms have been targeted by a band of hackers, the cybersecurity firm FireEye of Milpitas, California, announced in a report released on 1 December. Because stocks in these companies can gain or lose value rapidly on the back of events such as the release of clinical-trial results, access to their confidential business information can be particularly lucrative. The group of hackers, which FireEye has named 'Fin4', gains access to e-mail accounts and documents using personalized phishing messages. Of the more than 100 publicly traded businesses targeted by Fin4, 68% were health-care or pharmaceutical companies.

➔ NATURE.COM

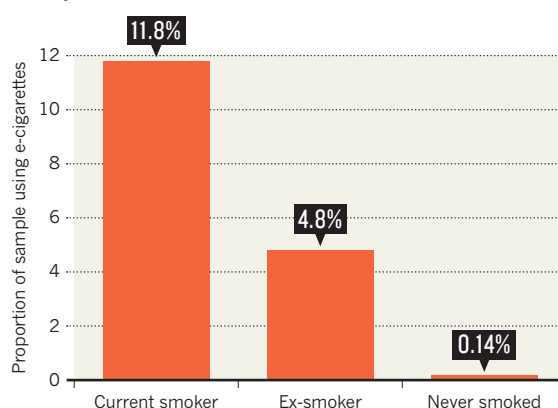
For daily news updates see:
www.nature.com/news

TREND WATCH

Only a tiny fraction of electronic-cigarette users in the United Kingdom are using the devices as their first introduction to nicotine, according to a report by the Office for National Statistics (ONS). Researchers are seeking evidence for whether people who have never smoked are tempted by the products — which provide nicotine by vaporizing a liquid, rather than burning tobacco. The ONS data (see chart) suggest that e-cigarettes are mainly used as an aid to quit smoking.

NON-SMOKERS RARELY VAPE

A UK survey suggests that electronic cigarettes are almost exclusively used by smokers and ex-smokers.



NEWS IN FOCUS

BIOENGINEERING Investigations launched into artificial-trachea work **p.16**



PALAEONTOLOGY Rival fossils put 'first bird' in its place **p.18**

MEDICINE Reports on off-label drug use to be mined for cancer clues **p.21**

ENERGY Scientists debate extent of US natural-gas reserves **p.28**

JOHN MOORE/GETTY



Testing for fever is one way to identify Ebola, but high temperatures can also be caused by a vaccine.

INFECTIOUS DISEASE

Ebola vaccine race begins

With trials under way, scientists are working out how to give vaccines in affected regions.

BY EWEN CALLAWAY

Safety trials of Ebola vaccines are starting to return results: at least one is known to be safe and to summon an immune response against the virus. The challenge now is to use the results to guide the larger studies that will reveal whether the vaccines work.

"The immune responses are there," says infectious-disease researcher Adrian Hill, director of the Jenner Institute in Oxford, UK. "The tough call is whether they're enough to protect humans against Ebola."

Published on 26 November (J. E. Ledgerwood *et al.* *N. Engl. J. Med.* <http://doi.org/xdr>; 2014), the first vaccine results to be released since the epidemic began were reported by researchers at the US National Institute of Allergy and Infectious Diseases (NIAID) in Bethesda, Maryland. The vaccine was developed by the NIAID and

the drug company GlaxoSmithKline (GSK), and is 'bivalent': it is made from a chimpanzee adenovirus that has been engineered to express proteins from two species of Ebola virus, known as Zaire and Sudan.

In phase 1 safety tests in September, 20 healthy volunteers in the United States received either a high or a low dose of the vaccine. Everyone mounted some kind of immune response against Ebola. But those who got the high dose produced more antibodies against a strain that is very similar to the virus driving the outbreak in West Africa; they also produced more immune cells known as CD8⁺ T cells.

Such safety studies are carried out in people who have a very low risk of becoming infected by Ebola so cannot reveal whether the vaccine protects against infection.

But Anthony Fauci, director of the NIAID, notes that the immune responses of the people who received the high dose were in line with those of monkeys that were protected from Ebola after receiving the same vaccine.

Others are taking the results with a pinch of salt, because the number of participants was low and because the vaccine used differs from the 'monovalent' vaccines that will be deployed in West Africa, which target only the Zaire species. "The big caveat here is this is not the vaccine that will be used in West Africa," says Hill.

Hill's team started administering a monovalent version of the vaccine to 60 UK volunteers in September. Volunteers in the United States, Mali and Switzerland have also received that vaccine as part of safety trials (Ebola arrived in Mali only after the trials began). Researchers will scrutinize the results of the studies to identify the best ►

► **NATURE.COM**
For more on Ebola, visit:
nature.com/ebola

► vaccines and the appropriate doses, as well as side effects, and use those insights to guide efficacy trials to be carried out in affected countries next year. “Next month we’ll have information on all of the candidates,” Hill says.

One issue is how to interpret the effects of the vaccines. Conventional wisdom says that having high levels of antibodies is the best sign of immunity. But animal studies have hinted that CD8⁺ T cells could be more important. A vaccine that raises both would be ideal, but if either remains low, developers are likely to add a booster in efficacy trials (some safety trials are already testing this approach). Data from the ongoing trials in Mali will be especially important to working out the best dose, because vaccines often generate weaker immune responses in sub-Saharan African populations than in others. This is partly because of the presence of malaria, which can suppress the immune system.

Testing vaccinated monkeys often involves exposing them to much higher levels of Ebola than humans typically encounter, so scientists may be overestimating the immune reaction needed to prevent infection, says Daniel Bausch, a physician at Tulane University School of Public Health and Tropical Medicine in New Orleans, Louisiana.

There are also side effects to consider. Two of the ten people who got the high dose developed a brief fever. Some researchers worry that vaccines that cause symptoms

“Next month we’ll have information on all of the candidates.”

that overlap with those of Ebola, such as fever, could sow confusion. But based on past studies of other adenovirus vaccines, Hill does

not expect such side effects to be common and says that they will be manageable, if people who get the vaccine are warned that they may develop.

The GSK/NIAID vaccines are not the only candidates set to deliver data. A vaccine made from a virus that infects livestock is being tested in the United States and Gabon, and soon in Kenya. The vaccine was developed by the Public Health Agency of Canada and licensed to NewLink Genetics in Ames, Iowa, and to Merck in Whitehouse Station, New Jersey. Phase 1 trials are to begin early next year for a regimen involving an adenovirus-based vaccine developed by the US pharmaceutical company Johnson & Johnson and the NIAID, and a booster made from vaccinia (a virus similar to the one that causes cowpox) by Bavarian Nordic in Denmark.

“What you have is something close to a race,” says Hill, “and the best ones should go forward.” ■



Surgeon Paolo Macchiarini carrying out the first transplant of a bioengineered synthetic trachea in 2011.

BIOENGINEERING

Artificial tracheas under scrutiny

The Karolinska Institute is carrying out two inquiries into an experimental transplant procedure.

BY DAVID CYRANOSKI

One of Europe’s most prestigious medical universities, the Karolinska Institute in Stockholm, has launched two investigations into the clinical procedures of a doctor famed for performing potentially revolutionary, bioengineered trachea transplants.

Since 2008, Paolo Macchiarini, a thoracic surgeon at the Karolinska Institute, has replaced parts of airways damaged by injury, cancer or other disorders in 17 patients. In the earlier cases, he transplanted parts of tracheas taken from cadavers; in his later work, he transplanted synthetic tracheas. In both procedures, before transplantation, he would treat the tracheas with stem cells taken from the patient’s bone marrow, which he says helps the transplants to act like biological tissue.

Bioengineering experts contacted by *Nature* say that Macchiarini’s procedures were considered a great leap for their nascent field because tracheas demand a high level of biological function — including the ability to defend against the constant assault of inhaled bacteria and to form a seal with the adjoining airway tissue.

Macchiarini’s reports were a “bright

spot” for the field, says David Mooney, a bioengineering specialist at Harvard University in Cambridge, Massachusetts.

One of the investigations is being conducted by an external expert in the relevant fields, who is due to report the findings on 15 January. It focuses on the three procedures that Macchiarini carried out at the Karolinska Institute, all of which involve artificial tracheas.

The investigation comes in response to a report filed in August by four thoracic doctors at the affiliated Karolinska Hospital — Matthias Corbascio, Thomas Fux, Karl-Henrik Grinnemo and Oscar Simonson — who helped to treat the three patients.

The doctors compared results in a paper Macchiarini published in *The Lancet* (*Lancet* 378, 1997–2004; 2011), describing the first use of a synthetic trachea seeded with stem cells, with the medical records of the patient. According to the doctors, there are discrepancies.

For example, the *Lancet* paper says that the synthetic trachea was “partly covered by nearly healthy epithelium”, indicating the growth of a protective cell layer, whereas the doctors say they could find no evidence in biopsy reports for healthy growth. The paper also says that the implanted trachea showed

KAROLINSKA UNIV. HOSPITAL VIA SCANPIX/CORBIS/REUTERS

signs of tight connection with the adjacent tissue and that it was acting like “an almost normal airway”, whereas bronchoscopy reports note gaps between the trachea and the bronchus, the tube that leads from the trachea to the lungs, and the need for stents to stabilize the airway.

“The problems alluded to are irreconcilable with the published reports,” says one US-based thoracic surgeon who reviewed the report but asked not to be named.

Macchiarini says that he will not reply to specific questions about alleged discrepancies between his publications and the medical records yet, because the allegations against him “now have to be investigated by an external expert, which is the normal process following cases of accusations of scientific misconduct”. He adds: “I certainly do welcome that investigation.” He is confident that “there is nothing suspect, unethical, inflated or misleading about anything I have done or reported”.

The Lancet says: “At this stage, we can’t comment on the allegations regarding Dr Macchiarini’s procedures.”

The complaint filed by the four doctors also claims that there were no informed-consent forms in the medical records for two of the three procedures carried out at the Karolinska Hospital. The one form on record was signed 17 days after the procedure.

Macchiarini says: “Of course there was consent. We would never have proceeded with the transplants if there wasn’t.” He adds: “I do not know why the form is dated 17 days after the procedure and can only assume it is some kind of clerical issue.” The patient “signed it in my presence, prior to the operation”. He adds that “there was absolutely no ethical breach”.

INVESTIGATION UNDER WAY

The Karolinska Institute’s ethics council, meanwhile, is carrying out the second investigation, which was launched in response to a report it received in June from Pierre Delaere, a head and neck surgeon at the University Hospital, KU Leuven, Belgium. Delaere complains, for example, that published descriptions of the transplants minimize complications faced by patients, such as the need for stents.

In August, Macchiarini sent a 15-page response to the Karolinska, acknowledging that he had “shortened” discussion of complications because “of the editor’s requirements during the review process”. But overall, he maintains that “all aspects of the patients’ outcomes are discussed in detail”. In an accompanying letter, Macchiarini says that he has “thoroughly reviewed” Delaere’s allegations and believes that they are “unfounded”.

The ethics council plans to interview the concerned parties during January, and to give recommendations to the vice-chancellor of the Karolinska Institute, Anders Hamsten, by the end of February at the earliest. Hamsten

will then decide how to proceed.

Hamsten says that the institute had initiated investigations into Macchiarini’s publications twice in the past, following allegations of scientific misconduct from other complainants. Both investigations concluded — one in July 2013, the other in September 2014 — that there was no scientific misconduct.

Macchiarini, who is currently spending part of his time at the Kuban State Medical University in Krasnodar, Russia, leading a project into the regeneration of lung airways, recently put on hold his own trial there into the use of synthetic tracheas after the death of a patient on 20 September 2014. The patient had had a synthetic transplant in June 2012 and, when that one began to fail, another in August 2013.

Macchiarini told *Nature* that the patient’s doctor has now reported the cause of death as “bilateral acute pneumonia with heart–lung insufficiency”, which he says is unrelated to the trachea transplant. He says that she was “breathing normally and asymptomatic” two weeks before her death. “We will be considering the restarting of the clinical trial now that this cause has been ascertained,” he told *Nature*.

The most recent of Macchiarini’s total of eight synthetic trachea-transplant patients, who was operated on in June, “is doing very well, is asymptomatic”, he says.

Six have died, with their post-transplant life-spans ranging from 3 to 31 months. Macchiarini says that one died because of complications following an accident, another from drinking too much alcohol, and another from “respiratory failure and subsequent multi-system organ failure”. In none, he says, has the death been linked to the transplant.

The remaining patient has been in intensive care ever since her procedure, more than two years ago. Macchiarini says this is not a result of the procedure. “When this patient came to Karolinska, her situation was dire,” he says. “Her doctors gave her a life expectancy of 3 to 6 months.” He also says that the surgery revealed more extensive damage to her airways than had been apparent from the examination before the operation. The damage could not have been diagnosed before the surgery, he says.

Nine other patients have received tracheas from cadavers. According to a paper by Macchiarini this year, four of those patients have died, either from recurrence of tumours or from gastrointestinal bleeding. Of the five still living, the paper reports that four are dependent on stents, and one has no need for stents (P. Jungebluth and P. Macchiarini *Thorac. Surg. Clin.* 24, 97–106; 2014).

Macchiarini emphasizes that the procedure is experimental. “Given the nature of this work, we are not in a position to guarantee them long-term survival and they are all abundantly aware of that going in,” he says. “We at least give them a chance, a chance at a longer life, and the hope of being the patient who survives long-term.” ■

Rival species recast significance of ‘first bird’

Archaeopteryx’s status is changing, but the animal is still key to the dinosaur–bird transition.

BY EWEN CALLAWAY

The iconic status of *Archaeopteryx*, the first animal discovered with both bird and dinosaur features, is under attack. More-recently discovered rival species show a similar mix of traits. But *Archaeopteryx* still hogged the opening symposium at the 2014 Society of Vertebrate Paleontology meeting in Berlin last month, and even festooned the official conference beer glasses.

As the simplistic idea of a ‘first’ bird gives way to a messy evolutionary transition, newly discovered fossils and improved analysis techniques put *Archaeopteryx* in prime position to unravel the details. “Research on *Archaeopteryx* is really catching a new breath,” says palaeobiologist Martin Kundrat of Uppsala University in Sweden, who co-organized the symposium at the November meeting.

The first *Archaeopteryx* fossil specimens turned up in limestone quarries in Bavaria, southern Germany, in the early 1860s. Until recently, they were the only fossil specimens found to mix bird- and dinosaur-like features. On the one hand, they are small — the fossils show juvenile creatures about the size of a magpie, which as adults may have been raven-sized — and have broad feathered wings that look good for gliding; on the other, they have a jaw with sharp teeth, dinosaur-like claws and a bony tail. These features led to the idea of the first bird, and generations of scientists have treated the 145-million-year-old animal as a ‘transitional species’ — the key piece of evidence linking birds and dinosaurs (*Archaeopteryx* is Greek for ‘ancient feather’, whereas its German name, *Urvogel*, means ‘first bird’).

But starting in the 1990s, the unique status of *Archaeopteryx* faced a challenge from the discovery in China of other potential transitional species. Fossils of *Anchiornis huxleyi* and *Microaptor xui* reveal small-bodied creatures like *Archaeopteryx*, and they may have used their four wings to glide. Another, *Aurornis xui*, has legs, claws and a tail similar to those of *Archaeopteryx*, yet lived about 10 million years earlier, leading some to propose it as a better candidate for first bird (see ‘The fight for first bird’).

Many scientists now believe that *Archaeopteryx* is just another dinosaur. Others find this hard to swallow. “To some ornithologists this is a really big deal — *Archaeopteryx* is the first

bird,” says Gareth Dyke, a vertebrate palaeontologist at the University of Southampton, UK. “They’d rather cut off one of their legs than admit it has nothing to do with bird origins.”

New specimens of *Archaeopteryx* are also reinvigorating research. There are now 12 known fossils of the creature, and at the symposium Kundrat unveiled the first scientific description of the eighth to be discovered. Researchers lost track of the ‘phantom’ after it was found in Germany in the early 1990s, until a collector purchased it in 2009. It is now on loan to a Bavarian museum with an agreement that it will always be available for research.

Kundrat reported that the phantom lived more recently than other *Archaeopteryx* fossils, bolstering suggestions that there may be more than one *Archaeopteryx* species. And in July, an analysis¹ of the eleventh fossil showed that the animals boasted feather-covered legs, suggesting that the feathers evolved for purposes other than flight, such as display or insulation. Kundrat says that the twelfth specimen might soon emerge from a private collection.

Scientists are also developing techniques to analyse *Archaeopteryx*. Kundrat presented a three-dimensional (3D) scan of the rock-embedded bones of the phantom, which he made with help from scientists at the European Synchrotron Radiation Facility in Grenoble, France. “Now we have the first very detailed picture of the *Archaeopteryx* skull,” he says.

Palaeontologist Ryan Carney of Brown University in Providence, Rhode Island, presented a 3D X-ray-tomography body scan of an *Archaeopteryx* fossil known as Thermopolis and a surface scan of another fossil. (Carney sports a black feather tattoo as a nod to his part in the discovery² of pigment structures in *Archaeopteryx*’s feathers that suggest the bird was black.) Both analyses will help to model whether and how *Archaeopteryx* flew and to determine its place in the transition from dinosaurs to birds.

Revelations about *Archaeopteryx*’s brain are already informing its place in the evolutionary tree. Last year, evolutionary biologist Amy Balanoff of Stony Brook University in New York and her colleagues reported³ in *Nature* that the

brains of *Archaeopteryx* and a variety of dinosaurs related to it were, like modern birds, all relatively large for their body sizes. Because earlier dinosaurs did not fly as well as modern birds, if they flew at all, the finding suggests that the brain expansion occurred well before flight emerged. It is yet another clue that *Archaeopteryx* may not have been any more bird-like than other suggested transitional species. “It’s another piece of data that has taken away from the specialness of *Archaeopteryx*,” says Balanoff.

Balanoff is now investigating whether the *Archaeopteryx* brain shares finer anatomical details — inferred from skull shape and small impressions on it — with theropod dinosaurs, the lineage that gave rise to birds. Her team also reported³ the discovery of a structure called a wulst in the *Archaeopteryx* brain, which was previously found only in living birds. The wulst, involved in perception, might have helped the creature to process visual information during flight. But the question is whether it, too, is present in *Archaeopteryx*’s dinosaur kin.

Taken together, the findings suggest that there was no neat transition from dinosaur to bird, making the idea of a first bird seem increasingly arbitrary and prone to the caprices of an incomplete fossil record. “I love *Archaeopteryx*. I did a thesis on the animal,” says Nicholas Longrich, a palaeontologist at the University of Bath, UK. “But in some ways, we’re moving beyond *Archaeopteryx*. It’s becoming one more fossil in this really diverse family tree.”

Some think the debate over the transition will one day be moot. “The boundary between dinosaurs and birds in the future will completely disappear,” says Ulysse Lefevre, a palaeontologist at the University of Liège in Belgium, who discussed new fossils from China at the Berlin meeting that might further blur the line.

What is clear is that *Archaeopteryx* research is still key to understanding the dinosaur–bird family tree. “There are some controversies over whether *Archaeopteryx* is the first bird or not. I don’t think it matters at the moment,” says Kundrat. “We are bringing *Archaeopteryx* back to life through new quantifiable analysis.” ■

1. Foth, C., Tischlinger, H. & Rahut, O. W. M. *Nature* **511**, 79–82 (2014).
2. Carney, R. M. et al. *Nature Commun.* **3**, 637 (2012).
3. Balanoff, A. M., Bever, G. S., Rowe, T. B. & Norell, M. A. *Nature* **501**, 93–96 (2013).

ILLUSTRATION BY EMILY WILLOUGHBY; GRAPHICS: JASIEK KRZYŻOPIAK/NATURE. SOURCES: BRAINS, REF/ TREE, J. CLARKE. SCIENCE **340**, 690–692 (2013); R. B. J. BENSON ET AL. *PLoS Biol.* **12**, e1001896 (2014).

The fight for first bird

For more than a century, scientists looked to *Archaeopteryx* to understand the transition from dinosaurs to birds. But more-recently discovered fossils are threatening to strip *Archaeopteryx* of its 'first bird' title. One by one, the features that once made *Archaeopteryx* stand out as a transitional species are falling away.



Zanabazar junior



Archaeopteryx



Woodpecker

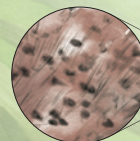
Zanabazar junior



BIRD BRAINS FOR EVERYONE

Despite their reputation, bird brains are big relative to the size of their bodies owing to an expanded cerebrum (red). *Archaeopteryx*'s brain was also large, but so were those of other theropod dinosaurs, such as *Zanabazar junior*, suggesting that this trait is not unique to birds.

Kulindadromeus zabaikalicus



Fossil of forelimb plates and bristles

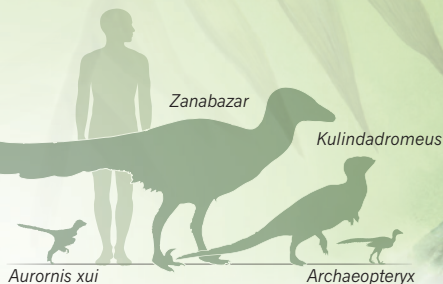
FALL OF FEATHERS

Feathers were once the sign of a bird. Things changed when feathered theropod dinosaurs — from the lineage that gave rise to modern birds — showed up in the 1990s. Now many researchers believe that all dinosaurs had feathers, thanks to *Kulindadromeus zabaikalicus*, which is from a completely different lineage. Its feathers — fossilized as structures that resemble ribbons, plates and bristles — may have been used for warmth or signalling.

ANATOMY ARGUMENT

Aurornis xui is the latest creature to challenge *Archaeopteryx*'s spot as first bird. As well as being of similar size, it has limbs, claws and a tail like *Archaeopteryx*, and it pre-dates its rival by about 10 million years.

Aurornis xui



FAMILY TREE

Birds are the dinosaurs that didn't go extinct. An incomplete fossil record for the transition means that each new fossil discovery can reshuffle the evolutionary order. Watch this space.





An untested device that sprays sea water into the stratosphere is one futuristic idea for cooling the planet.

GEOENGINEERING

Climate tinkerers thrash out a plan

Geoengineers meet to work out what research is acceptable.

BY QUIRIN SCHIERMEIER

On 1 December, the United Nations kicked off a summit in Lima that aims to forge a global deal to reduce greenhouse-gas emissions. Now, representatives of dozens of scientific societies are gathering in Washington DC to thrash out a set of principles for researching highly controversial technologies known as geoengineering. The methods offer ways to cool the planet should political approaches fail.

"There are a number of risks and unknowns," says Paul Bertsch, deputy director of the Land and Water Flagship at the Commonwealth Scientific and Industrial Research Organisation in Brisbane, Australia,

and past chair of the Council of Scientific Society Presidents, which is convening the geoengineering meeting. "So we urgently need to develop and implement a coordinated research plan that begins to address these in a deliberate way."

Some ideas, such as injecting carbon dioxide into rocks or the depths of the ocean, are already being tested. Others are more futuristic: spraying sea water into the air to brighten clouds and reflect more sunlight back into space; adding sulphate particles to the upper atmosphere to mimic the natural cooling effect of volcanic ash; and even placing giant mirrors into orbit to reflect sunlight before it reaches Earth.

Not one, however, has garnered much enthusiasm in environmental or political

spheres. The idea of tinkering with the planet smacks of scientific hubris, and many are worried about unintended consequences. Climate scientists are concerned, for example, that adding sulphate to the stratosphere might reduce rainfall in some regions and worsen ozone depletion.

S. SALTER/J. MACNEILL

On 2–3 December, leaders of societies representing some 1.4 million scientists, engineers and educators will work out what research is and is not acceptable given the possible social, ecological and economic effects of climate engineering. A conference held in 2010 in Asilomar, California, failed to produce clear guidelines (see *Nature* **464**, 656; 2010).

Most scientists say that it is too early to consider large-scale trials, especially for solar-radiation management, because the techniques have not yet been adequately tested in controlled settings. However, many maintain that geoengineering should not be ruled out as a last resort to prevent the worst effects of global warming.

"The question is when, if at all, should we start doing outdoor experiments?" says Matthew Watson, a volcanologist at the University of Bristol, UK, who is overseeing a project to determine how the deliberate spreading of sun-blocking particles might alter atmospheric chemistry (see 'UK experiments'). "I don't particularly 'like' geoengineering, but I'm afraid we do need to think about controlled field trials."

A method that has already been tested — ocean fertilization — provides a particularly thorny case study. The idea was to boost ocean uptake of carbon dioxide by pouring iron into the sea to stimulate the growth of algal blooms. When the algae die, the captured carbon sinks to the ocean floor, where it may remain locked away for centuries.

But the approach came under fire when entrepreneurs smelled business opportunities. Plans by companies in the United States and Australia to fertilize large swathes of ocean to generate carbon credits that could be sold on greenhouse-gas-emissions markets were headed off by a 2008 amendment to the London Convention, an international treaty that governs ocean pollution.

Together with a resolution made under the United Nations Convention on Biological Diversity a few months earlier, the amendment made it difficult to conduct trials of ocean fertilization. In 2009, for example, an international research cruise was stopped en route to the Southern Ocean over fears that an iron-stimulated algal bloom the team had planned to encourage there might violate international law.

Meanwhile, another attempt, by an amateur scientist in 2012 off the coast of British Columbia, led to an international storm of protest and prompted heated discussions in the Canadian government over the legality of the experiment.

Such unresolved governance issues mean

UK EXPERIMENTS

Results raise questions

The teams behind three UK geoengineering studies have reported decidedly mixed results.

Matthew Watson, a volcanologist at the University of Bristol, presented the results of the Stratospheric Particle Injection for Climate Engineering (SPICE) project. SPICE investigated whether spraying particles into the atmosphere could reflect sunlight and cool the planet, offsetting global warming. A planned test of some of the technology was abandoned in 2012 when conflict-of-interest issues emerged over a patent application for the system. But Watson says that SPICE produced useful insights, such as how a large-scale project might alter the Sahel region in Africa.

Piers Forster at the University of Leeds, who led the Integrated Assessment of Geoengineering Proposals project, said that his team's computer modelling showed that several techniques to manage the Sun's radiation would produce damaging changes in rainfall that could affect 25–65% of the world's population.

Watson, Forster and the University of Oxford's Steve Rayner, who is leader of a third effort called the Climate Geoengineering Governance project, agreed that their work created many questions. **Daniel Cressey**

that little funding is available for further studies. "We're caught up in politics," says Ken Buesseler, an ocean scientist at Woods Hole Oceanographic Institution in Massachusetts. "You'd absolutely like to avoid rogue experiments that don't generate proper science. But there is every reason to pursue real science in the field in an open and responsible way."

Meeting discussions are aimed at creating comprehensive guidelines for the safe conduct of field experiments.

Neither ocean fertilization nor any other single activity will solve the global warming problem, cautions Anya Waite of the Alfred Wegener Institute in Bremerhaven, Germany, who represents the fields of oceanography and limnology at this week's meeting. "But limited ocean-fertilization experiments are telling us a lot about how biological processes in the ocean control climate. In terms of new regulations, they should be the first cab off the ranks." ■ **SEE EDITORIAL P.8**

ESA/PLANCK COLLABORATION



SUZANNE KREITER/THE BOSTON GLOBE/GETTY

People with cancer often receive drugs that are not yet approved for their specific diagnosis.

MEDICAL RESEARCH

Projects seek hidden effects of cancer drugs

Researchers gather data on innovative uses of cancer treatments.

BY HEIDI LEDFORD

More than 60% of US prescriptions for cancer drugs call for using the medications in ways that are not approved by the government. Often, they are the only hope against a fatal illness that has thwarted conventional treatments. But although such 'off-label' use is common, it is hard to know

how effective it is because outcomes are not being tracked systematically.

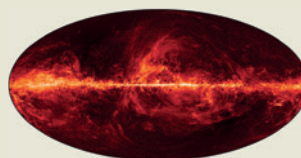
Thousands of small-scale experiments are going on all the time in clinics and hospitals — but, says Amy Abernethy, an oncologist at Duke University in Durham, North Carolina, "we haven't been able to learn from them".

On 21 November, researchers gathering in Washington DC for the annual Conference ►



**MORE
ONLINE**

TOP NEWS



Planck probe dashes dark-matter hopes go.nature.com/vgszc3

MORE NEWS

- How humans see infrared light go.nature.com/slblpx
- Graphene makes sieve for hydrogen fuel cells go.nature.com/cpxy6h
- How EU budget impasse imperils science go.nature.com/wfaadh

► on Clinical Cancer Research discussed two programmes, both planned for launch next year, that will aim to capture data from a subset of those experiments. The intent is to provide databases for researchers to mine — looking not just for potentially effective treatments, but also for cellular clues as to why individual cancers are resistant or susceptible to drugs.

“It’s going to be what we learn from our patients that leads us back to understanding the cancer biology,” says Vincent Miller, chief medical officer of Foundation Medicine, a company in Cambridge, Massachusetts, that provides genetic profiles of tumours.

Although it is legal for physicians to prescribe drugs off-label in the United States, such use is often controversial, and drug-makers pay steep fines if they promote their products for unapproved uses.

STANDARD PRACTICE

Off-label drug regimens are woven into the fabric of cancer care and are expected to become more prevalent as treatments become more tailored to the genetics of individual tumours. But genetics is not everything: drugs can be approved only for the cancers in which they have been tested, even if the mutation that they target is shared by several tumour types. Vemurafenib and dabrafenib, for example,

target a mutation in a protein called BRAF. They have been approved to treat melanoma, but not lung or thyroid cancers that have the same mutation.

The problem is partly one of numbers, says Richard Schilsky, chief medical officer of the American Society of Clinical Oncology (ASCO) in Alexandria, Virginia. “There are not

“It’s going to be what we learn from our patients that leads us back to understanding the cancer biology.”

enough patients and not enough money to test every drug in every subtype of cancer using a randomized clinical trial,” he says. To address this shortcoming, Schilsky is spearheading a programme called the TAPUR (Targeted Agent and Profiling Utilization Registry) Study, to be launched in mid-2015. The study, run through ASCO, will compile information about treatments, outcomes and mutations in people who have exhausted conventional therapies and moved on to unapproved treatments.

A similar effort, led by Dane Dickson, director of clinical science for molecular diagnostics at the health benefits provider Palmetto of Columbia, South Carolina, has brought together pharmaceutical companies, researchers and patient advocates to launch

another registry that aims to capture data about off-label use. The programme, called MED-C (Molecular Evidence Development Consortium), will require every person who enrolls to undergo a standardized genetic test so that researchers can better compare results from different hospitals.

Both programmes are in discussions with pharmaceutical companies to help to provide therapies free-of-cost to participants. This is intended to motivate physicians to participate. In exchange, the companies have an opportunity to gather more data about their drugs.

The benefit for patients is clear, says Ellen Sigal, chairwoman and founder of Friends of Cancer Research, an advocacy group based in Washington DC, which co-hosted the conference with the Brookings Institution, a nearby think-tank. “Off-label use is happening and we’d better figure out a way to get the data and do it in a meaningful way,” she says. “Right now, we are getting nothing.” ■

CORRECTION

The World View ‘Open access is tiring out peer reviewers’ (*Nature* **515**, 467; 2014), erroneously calculated the percentage rise in the number of articles indexed in Scopus as 213%. In fact, the increase is 113%.



QUANTUM COMPUTER QUEST

After a 30-year struggle to harness quantum weirdness for computing, physicists finally have their goal in reach.

BY ELIZABETH GIBNEY

When asked what he likes best about working for Google, physicist John Martinis does not mention the famous massage chairs in the hallways, or the free snacks available just about anywhere at the company's campus in Mountain View, California. Instead, he marvels at Google's tolerance of failure in pursuit of a visionary goal. "If every project they try works," he says, "they think they aren't trying hard enough."

Martinis reckons that he is going to need that kind of patience. In September, Google recruited him and his 20-member research team from the University of California, Santa Barbara, and set them to work on the notoriously difficult task of building quantum computers: devices that exploit the quirks of the quantum world to carry out calculations that ordinary computers could not finish in the lifetime of the Universe.

It is a vision that has frustrated Martinis and many other physicists ever since it was proposed in the early 1980s. In practice, the quantum effects essential in such a computer are incredibly fragile and hard to control: if one stray photon or vibration from the outside hits the device in the wrong way, the calculation will collapse. Even today, after three decades of effort, the best quantum computers in the world are barely able to do school-level problems such as finding the prime factors of the number 21. (Answer: 3 and 7.)

The result has been a rate of progress so slow that sceptics often compare quantum computing to fusion energy: it is a revolutionary technology that always seems to be decades away.

But maybe not. Many physicists in the field think that their 30-year slog may finally be on the verge of paying dividends. Not only can they now generate quantum bits, or 'qubits', that last for minutes instead of nanoseconds, they are also much better at correcting the system when errors arise from outside perturbations and other causes. At the same time, quantum-software engineers are coming up with applications that could justify the expense of developing these machines, such as finding new catalysts for industrial processes.

The prospects for useful and profitable quantum computers are good enough to have drawn Google into the game, along with IBM and Microsoft, among others. Several academic groups are also pushing the technology in practical directions. At the Delft University of Technology in the Netherlands, for example, the government-backed QuTech Centre is bringing researchers together with the Dutch high-tech industry. Delft physicist Ronald Hanson says that he will be able to make the building blocks of a universal quantum computer in just five years, and a fully functional — if bulky and inefficient — demonstration machine in a little more than a decade.

Martinis says that he has no fixed timetable, but is just as optimistic. "We got a lot of things working in the last couple of years," he says. "It is still possible that nature just won't allow it to work, but I think we have a decent chance."

SEVENTIES CHILD

The conceptual foundations of quantum computing were laid during the 1970s and early 1980s — most notably by the late US physicist Richard Feynman, whose lecture on the subject, published¹ in 1982, is widely credited with launching the field. The basic insight is that conventional computers are 'either-or' machines, meaning that the tiny silicon circuit that encodes a given bit of information acts like a switch that is either open or closed. This means that it can represent choices such as 'true' or 'false', or the 1s and 0s of binary arithmetic. But in the quantum realm, 'either-or' gives way to 'both-and': if binary 1s are represented by, say, electrons that are spinning clockwise, and 0s by electrons spinning counterclockwise, then the subatomic laws that govern those particles make it possible for a given quantum bit to be both 1 and 0 at the same time.

By extension, the set of qubits comprising the memory of a quantum computer could exist in every possible combination of 1s and 0s at once. Where a classical computer has to try each combination in turn, a quantum computer could process all those combinations simultaneously — in effect, carrying out calculations on every possible set of input data in parallel. And because the number of combinations increases exponentially with the size of the memory, the quantum computer has the potential to be exponentially faster than its classical counterpart.

That insight became much more than a scientific curiosity in 1994, when the US mathematician Peter Shor developed an algorithm that would allow a quantum computer to factor large numbers very quickly². Such factorization is

prohibitively time-consuming for standard computers, which is why it forms the basis for widely used encryption techniques. Shor's algorithm meant that in principle, quantum computers could crack that encryption.

Then two years later, Lov Grover, a researcher at Bell Labs in Murray Hill, New Jersey, devised another algorithm that showed how quantum computers could radically speed up searches of massive databases³.

The demonstration of such obviously important applications quickly attracted researchers and funding — accompanied by claims that working quantum computers would be ready in a matter of years. "But in hindsight they were naive," says Hanson. Researchers have been able to make some progress by devising special-purpose quantum devices that are tailored for solving specific problems (see *Nature*

491, 322–324; 2012 and *Nature* **498**, 286–288; 2013). But achieving the ultimate goal — a general-purpose, digital quantum computer that can be programmed to carry out any algorithm — has proved much tougher.

The problem is the extreme fragility of quantum effects: any slight influence from the outside world will cause a qubit to collapse so that it no longer represents many different states at once. If qubits are going to be useful in real-world calculations, they must be kept in

the strictest isolation and manipulated with care — extremely difficult tasks. They also need to remain in their quantum states for much longer than it takes to perform a computing step — typically a microsecond or so.

To achieve those goals, physicists are pursuing a two-fold strategy: extending the life of qubits and reducing how often they go wrong, and devising algorithms that can correct any errors that do occur.

The qubit design currently favoured by many researchers is based on microchip-scale circuits made from superconductors, materials that lose all resistance to the flow of electricity at very low temperatures. Thanks to a quantum phenomenon known as the Josephson effect, electric currents flowing around tiny loops in such circuits can circle both clockwise and counterclockwise at once, so are perfect for representing a qubit. Such circuits are tricky to implement, says Martinis. "You have to work many years to figure out all the physics." But after a decade spent refining designs and learning how to isolate the circuits from the environment, his group and others have increased qubit lifetimes by a factor of 10,000, meaning that they can now regularly maintain their state for around 50 to 100 microseconds. They have also slashed the rate at which errors occur by finding better ways to manipulate and control their qubits as the computation proceeds.

Lifetimes have been tougher to boost for qubits that are based on the spins of electrons or atomic nuclei, because these spins are easily flipped by the magnetic fields of neighbouring particles. In October, however, Andrea Morello and Andrew Dzurak, physicists at the University of New South Wales in Sydney, Australia, announced⁴ that they had eliminated such interference by embedding spin-qubits in purified silicon that contains no magnetic isotopes of the element. The resulting qubits lived as long as 30 seconds.

In 1997, physicist Alexei Kitaev of the California Institute of Technology in Pasadena proposed⁵ a more radical approach: make qubits out of anyons, which are states of matter that arise from the collective properties of many particles, yet behave as just one particle. Some anyons have another special property: their quantum state reveals a history of their recent interactions. If these anyons were used as qubits, Kitaev argued, the order of their interactions could encode information. And because this encoding is effectively spread throughout the system, the qubits would have a natural protection against errors arising in any individual part.

Known as 'topological qubits', these entities remain theoretical, but the idea shows enough promise that Microsoft and a number of other companies are investing in efforts to create them in the laboratory.

Even with the most robust qubits, however, errors are inevitable. That is also the case in ordinary computers, but errors are particularly

**"THERE ARE NO
FUNDAMENTAL
ROADBLOCKS LEFT."**

➔ **NATURE.COM**

To hear about
quantum-computer
development, see:
go.nature.com/b2jda4

troublesome in a quantum computer because they grow exponentially with the number of qubits. “One of the real tricks of eventually building a quantum computer is finding a way to get around that,” says David Cory, an experimental quantum physicist at the University of Waterloo in Canada.

That means implementing some form of quantum error correction. In standard computers, correcting for errors can be as simple as starting off with multiple copies of each bit. A majority vote among the copies can reveal whether any one of them has later flipped from a 1 to 0 or vice versa. That does not work in the quantum world because it is impossible to copy a qubit without destroying its quantum state. But qubits can be compared, so theorists have tried to devise correction schemes that ask various pairs of qubits whether they have the same or different values, and then use the answers to deduce whether individual qubits have gone wrong.

Until recently, a big problem was that qubits typically made about one error in every ten computer steps, and the available correction schemes could not begin to keep up. “Theorists were saying we need average error rates to be, say, 1 in 100,000 operations,” says John Morton, an experimental physicist at University College London. In April this year, however, Martinis and his group announced⁶ that they had demonstrated a ‘surface-code’ scheme that spreads the quantum information of a qubit among several physical qubits, similar to what Kitaev proposed for topological qubits. In its publication, the group described how it had used this technique to implement 5 qubits of information in a way that could handle error rates as high as 1 per 100 operations — a rate that they and others⁷ are now able to achieve (see page 10).

ONWARDS AND UPWARDS

Together, improvements in qubit error rates and the ability of codes to cope with errors have radically changed the outlook of the field, says Morton. “What makes it an exciting time is that we can now focus on scaling up,” he says.

At the QuTech Centre, Hanson agrees. “There are no fundamental roadblocks left,” he says. He is now advertising for 5 electrical engineer professorships, and looking for 40 technicians and researchers, so that he can scale up from laboratory experiments to practical technology. Their main tasks will be to figure out how to fabricate large-scale qubit arrays, how to control the quantum computation and read out the results and how to connect up the quantum circuitry to classical electronics that reside on the same chip.

Both Hanson and his colleague Lieven Vandersypen, who leads Delft’s efforts to develop spin qubits embedded in the tiny semiconductor crystals known as quantum dots, aim to build arrays of 17 qubits in the next 5 years. This, they say, is the minimum to demonstrate that the surface-code scheme works as hoped. To create a single virtual qubit that remains correct over the hours it takes to run real algorithms may mean spreading its information over 100 physical qubits. Each extra qubit increases the complexity of the hardware. But once a team has acquired the know-how to create a few dozen physical qubits, they believe, growing to the hundred they need to make a handful of virtual qubits should be much easier. “Then it’s a case of ambitious engineering to go to 100, or 1,000. I hope that in 10 years we’ll be talking about 100s of qubits,” says Vandersypen.

At the Swiss Federal Institute of Technology in Zurich, however, theoretical physicist Matthias Troyer cautions that the goal of hundreds of qubits will not be easy or cheap to achieve. Assuming that quantum chips will be as least as hard to manufacture as semiconductor chips, Troyer estimates that working out how to wire up, manipulate and fabricate qubits in bulk will be a US\$10-billion problem. That poses a crucial question, he says. “Why should one do it?”

Troyer has spent the past three years looking for an answer — a ‘killer app’ for quantum computing that would make the development costs worthwhile. The two classic examples, code-cracking and searching databases, are not good enough, says Troyer. Shor’s algorithm will require thousands of qubits to do any serious factorization, he says, and there are other forms of encryption that a quantum computer would do nothing to solve. And although quantum computers may search databases faster, they are still limited by the time it takes to feed the data into the circuit, which would not change.

Troyer thinks that a much more fruitful application for the near future is the modelling of electrons in materials and molecules — something that quickly becomes too difficult for today’s supercomputers. At first, this, too, seemed a long shot. His early estimates suggested that it would take a quantum computer as long as 300 years to simulate the molecular dynamics of even a small molecule — such as the iron sulphide inside the ferredoxin proteins that are involved in nitrogen fixation in plants. “Clearly, that was on the border of being science fiction,” he says. But by rewriting the software⁸, he brought the figure down

to 30 years — then to just 300 seconds. “Just like in classical computing, where one has to sit down and optimize the algorithm,” he says, “the same is needed for a quantum algorithm.”

With around 400 encoded qubits, Troyer says, it would be possible to analyse ways to improve industrial nitrogen fixation — the energy-intensive process that turns the unreactive molecule in air into fertilizer. This reaction is now carried out on an industrial scale using the 116-year-old Haber process, but that uses up about 5% of the natural gas produced each year worldwide. Troyer thinks that a quantum computer could help to design a catalyst that would be much more energy-efficient than the current ones. “That would be worth building a quantum computer for,” he says.

Other killer applications might be searching for new high-temperature superconductors, or improving the catalysts used to capture carbon from the air or from industrial exhaust streams. “All these are important questions. If it makes progress there, easily that’s your 10 billion,” says Troyer.

For now, however, Martinis and other veterans of the field caution that quantum computing is still in the early stages. Although industry is now deep into the research, no one even has one of these things to play with. Quantum computing today is comparable to conventional computing in the years after the Second World War, he says, when every device was a laboratory experiment that had been crafted by hand. “We’re somewhere between the invention of the transistor and the invention of the integrated circuit,” he concludes. At Google, the project has the buzz of a Silicon Valley start-up, says Martinis, albeit one with hefty backing. After years of the hard work of perfecting qubits, he is happy to finally be able to focus on building a quantum computer that can actually solve real problems. “Google created a new name for scientists working on the hardware effort, ‘quantum engineers,’” says Martinis. “This is a dream job for me.” ■

Elizabeth Gibney is a reporter for Nature based in London.

1. Feynman, R. P. *Int. J. Theoret. Phys.* **21**, 467–488 (1982).
2. Shor, P. W. *Proc. 35th Ann. Symp. Found. Comp. Sci. IEEE* 124–134 (1994).
3. Grover, L. K. *Proc. 28th Ann. ACM Symp. Theory Comput.* 212–219 (1996).
4. Muhonen, J. T. *et al. Nature Nanotechnol.* <http://dx.doi.org/10.1038/nnano.2014.211> (2014).
5. Kitaev, A. Y. *Ann. Phys.* **303**, 2–30 (2003).
6. Barends, R. *et al. Nature* **508**, 500–503 (2014).
7. Hart, T. P. *et al. Phys. Rev. Lett.* **113**, 220501 (2014).
8. Poulin, D. *et al. Preprint at* <http://arxiv.org/abs/1406.4920> (2014).



THE FRACKING FALLACY

The United States is banking on decades of abundant natural gas to power its economic resurgence. That may be wishful thinking.

BY MASON INMAN

When US President Barack Obama talks about the future, he foresees a thriving US economy fuelled to a large degree by vast amounts of natural gas pouring from domestic wells. “We have a supply of natural gas that can last America nearly 100 years,” he declared in his 2012 State of the Union address.

Obama’s statement reflects an optimism that has permeated the United States. It is all thanks to fracking — or hydraulic fracturing — which has made it possible to coax natural gas at a relatively low price out of the fine-grained rock known as shale. Around the country, terms such as ‘shale revolution’ and ‘energy abundance’ echo through corporate boardrooms.

Companies are betting big on forecasts of cheap, plentiful natural gas. Over the next 20 years, US industry and electricity producers are expected to invest hundreds of billions of dollars in new plants that rely on natural gas. And billions more dollars are pouring into the construction of export facilities that will enable

the United States to ship liquefied natural gas to Europe, Asia and South America.

All of those investments are based on the expectation that US gas production will climb for decades, in line with the official forecasts by the US Energy Information Administration (EIA). As agency director Adam Sieminski put it last year: “For natural gas, the EIA has no doubt at all that production can continue to grow all the way out to 2040.”

But a careful examination of the assumptions behind such bullish forecasts suggests that they may be overly optimistic, in part because the government’s predictions rely on coarse-grained studies of major shale formations, or plays. Now, researchers are analysing those formations in much greater detail and are issuing more-conservative forecasts. They calculate that such formations have relatively small ‘sweet spots’ where it will be profitable to extract gas.

The results are “bad news”, says Tad Patzek, head of the University of Texas at Austin’s department of petroleum and geosystems engineering, and a member of the team that is conducting the in-depth analyses. With companies trying to extract shale gas as fast as possible and export significant quantities, he argues, “we’re setting ourselves up for a major fiasco”.

That could have repercussions well beyond the United States. If US natural-gas production falls, plans to export large amounts overseas could fizzle. And nations hoping to tap their own shale formations may reconsider. “If it begins to look as if it’s going to end in tears in the United States, that would certainly have an impact on the enthusiasm in different parts of the world,” says economist Paul Stevens of Chatham House, a London-based think tank.

The idea that natural gas will be abundant

A rig drills for natural gas using hydraulic-fracturing methods in a Pennsylvania shale formation.

JIM LO SCALZO/EPA/ALAMY

is a sharp turnaround from more pessimistic outlooks that prevailed until about five years ago. Throughout the 1990s, US natural-gas production had been stuck on a plateau. With gas supplying one-quarter of US energy, there were widespread worries that supplies would shrink and the nation would become dependent on imports. The EIA, which collects energy data and provides a long-term outlook for US energy, projected as recently as 2008 that US natural-gas production would remain fairly flat for the following couple of decades.

Then the shale boom caught everyone by surprise. It relied on fracking technology that had been around for decades — but when gas prices were low, the technology was considered too costly to use on shale. In the 2000s, however, prices rose high enough to prompt more companies to frack shale formations. Combined with new techniques for drilling long horizontal wells, this pushed US natural-gas production to an all-time high, allowing the nation to regain a title it had previously held for decades: the world's top natural-gas producer.

RICH ROCKS

Much of the credit for that goes to the Marcellus shale formation, which stretches across West Virginia, Pennsylvania and New York. Beneath thickly forested rolling hills, companies have sunk more than 8,000 wells over several years, and are adding about 100 more every month. Each well extends down for about 2 kilometres before veering sideways and snaking for more than a kilometre through the shale. The Marcellus now supplies 385 million cubic metres of gas per day, more than enough to supply half of the gas currently burned in US power plants.

A substantial portion of the rest of the US gas supply comes from three other shale plays — the Barnett in Texas, the Fayetteville in Arkansas and the Haynesville, which straddles the Louisiana–Texas border. Together, these 'big four' plays boast more than 30,000 wells and are responsible for two-thirds of current US shale-gas production.

The EIA — like nearly all other forecasters — did not see the boom coming, and has consistently underestimated how much gas would come from shale. But as the boom unfolded, the agency substantially raised its long-term expectations for shale gas. In its *Annual Energy Outlook 2014*, the 'reference case' scenario — based on the expectation that natural-gas prices will gradually rise, but remain relatively low — shows US production growing until 2040, driven by large increases in shale gas.

The EIA has not published its projections for individual shale-gas plays, but has released them to *Nature*. In the latest reference-case forecast, production from the big four plays would continue rising quickly until 2020, then plateau for at least 20 years. Other shale-gas plays would keep the boom going until 2040 (see 'Battle of the forecasts').

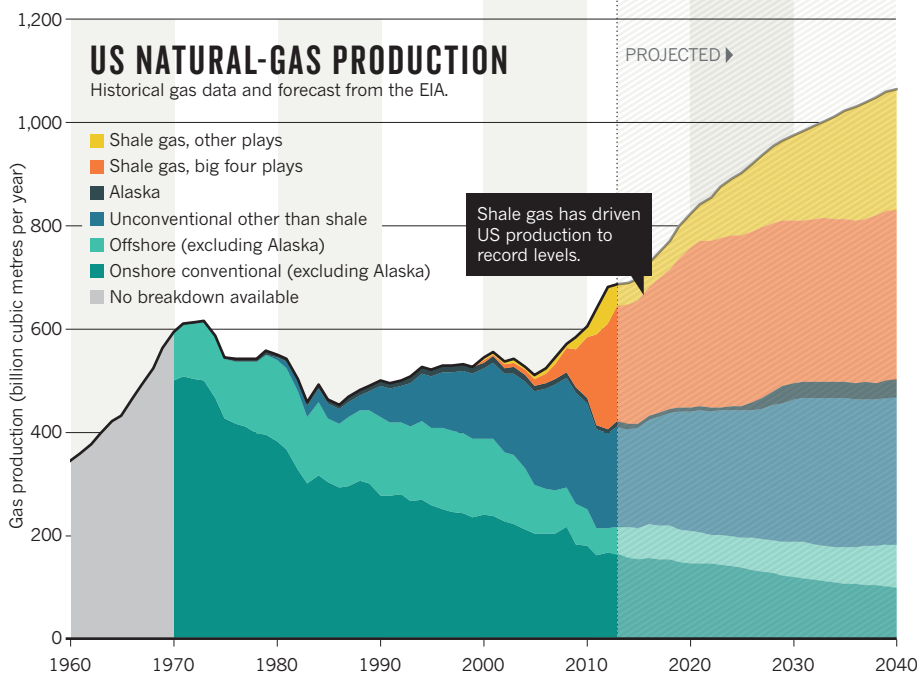
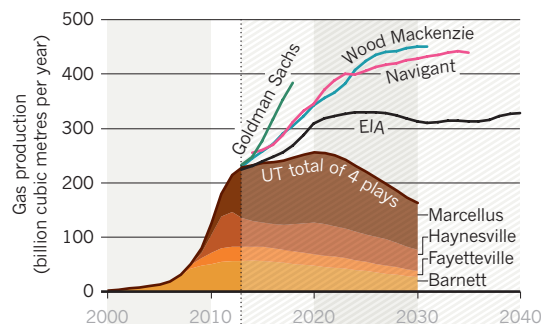
Petroleum-industry analysts create their

BATTLE OF THE FORECASTS

Production of natural gas in the United States is climbing rapidly, and the US Energy Information Administration (EIA) predicts long-term growth. But studies by the University of Texas (UT) challenge that forecast.

BIG FOUR SOURCES

The Texas team made forecasts for the four most productive shale-gas formations, or plays. Those forecasts suggest that gas production will peak soon and quickly drop, a much more pessimistic outlook than those offered by the EIA and several companies, such as Goldman Sachs.



own shale-gas forecasts, which generally fall in the neighbourhood of the EIA assessment. "EIA's outlook is pretty close to the consensus," says economist Guy Caruso of the Center for Strategic and International Studies in Washington DC, who is a former director of the agency. However, these consultancies rarely release the details behind their forecasts. That makes it difficult to assess and discuss their assumptions and methods, argues Ruud Weijermars, a geoscientist at Texas A&M University in College Station. Industry and consultancy studies are "entirely different from the peer-reviewed domain", he says.

To provide rigorous and transparent forecasts of shale-gas production, a team of a dozen geoscientists, petroleum engineers and economists at the University of Texas at Austin has spent more than three years on a systematic set of studies of the major shale plays. The research was funded by a US\$1.5-million grant from the Alfred P. Sloan Foundation in New York City, and has been appearing gradually in academic

journals^{1–5} and conference presentations. That work is the "most authoritative" in this area so far, says Weijermars.

If natural-gas prices were to follow the scenario that the EIA used in its 2014 annual report, the Texas team forecasts that production from the big four plays would peak in 2020, and decline from then on. By 2030, these plays would be producing only about half as much as in the EIA's reference case. Even the agency's most conservative scenarios seem to be higher than the Texas team's forecasts. "Obviously they do not agree very well with the EIA results," says Patzek.

The main difference between the Texas and EIA forecasts may come down to how fine-grained each assessment is. The EIA breaks up each shale play by county, calculating an average well productivity for that area. But counties often cover more than 1,000 square kilometres, large enough to hold thousands of horizontal fracked wells. The Texas team, by contrast, splits each play into blocks of one square mile

(2.6 square kilometres) — a resolution at least 20 times finer than the EIA's.

Resolution matters because each play has sweet spots that yield a lot of gas, and large areas where wells are less productive. Companies try to target the sweet spots first, so wells drilled in the future may be less productive than current ones. The EIA's model so far has assumed that future wells will be at least as productive as past wells in the same county. But this approach, Patzek argues, "leads to results that are way too optimistic".

The high resolution of the Texas studies allows their model to distinguish the sweet spots from the marginal areas. As a result, says study co-leader Scott Tinker, a geoscientist at the University of Texas at Austin, "we've been able to say, better than in the past, what a future well would look like".

The Texas and EIA studies also differ in how they estimate the total number of wells that could be economically drilled in each play. The EIA does not explicitly state that number, but its analysis seems to require more wells than the Texas assessment, which excludes areas where drilling would be difficult, such as under lakes or major cities. These features of the model were chosen to "mimic reality", Tinker says, and were based on team members' long experience in the petroleum industry.

ALTERNATIVE FUTURES

The lower forecasts from Texas mesh with a few independent studies that use simpler methods. Studies by Weijermars⁶, as well as Mark Kaiser⁷ of Louisiana State University in Baton Rouge and retired Geological Survey of Canada geologist David Hughes⁸, suggest that increasing production, as in the EIA's forecasts, would require a significant and sustained increase in drilling over the next 25 years, which may not be profitable.

Some industry insiders are impressed by the Texas assessment. Richard Nehring, an oil and gas analyst at Nehring Associates in Colorado Springs, Colorado, which operates a widely used database of oil and gas fields, says the team's approach is "how unconventional resource assessments should be done".

Patzek says that the EIA's method amounts to "educated guesswork". But he and others are reluctant to come down too hard. The EIA is doing "the best with the resources they have and the timelines they have", says Patzek. Its 2014 budget — which covers data collection and forecasting for all types of energy — totalled just \$117 million, about the cost of drilling a dozen wells in the Haynesville shale. The EIA is "good value for the money", says Caruso. "I always felt we were underfunded. The EIA was being asked to do more and more, with less and less."

Patzek acknowledges that forecasts of shale plays "are very, very difficult and uncertain", in part because the technologies and approaches to drilling are rapidly evolving. In newer plays, companies are still working out the best spots

to drill. And it is still unclear how tightly wells can be packed before they significantly interfere with each other.

Representatives of the EIA defend the agency's assessments and argue that they should not be compared with the Texas studies

"WE'RE SETTING OURSELVES UP FOR A MAJOR FIASCO."

because they use different assumptions and include many scenarios. "Both modelling efforts are valuable, and in many respects feed each other," says John Staub, leader of the EIA's team on oil and gas exploration and production analysis. "In fact, EIA has incorporated insights from the University of Texas team," he says.

Yet in a working paper⁹ published online on 14 October, two EIA analysts acknowledge problems with the agency's methods so far. They argue that it would be better to draw upon high-resolution geological maps, and they point to those generated by the Texas team as an example of how such models could improve forecasts by delineating sweet spots. The paper carries a disclaimer that the authors' views are not necessarily those of the EIA — but the agency does plan to use a new approach along these lines when it assesses the Marcellus play for its 2015 annual report. (When *Nature* asked the authors of that paper for an on-the-record interview, they referred questions to Staub.)

BOOM OR BUST

Members of the Texas team are still debating the implications of their own study. Tinker is relatively sanguine, arguing that the team's estimates are "conservative", so actual production could turn out to be higher. The big four shale-gas plays, he says, will yield "a pretty robust contribution of natural gas to the country for the next few decades. It's bought quite a bit of time."

Patzek argues that actual production could come out lower than the team's forecasts. He talks about it hitting a peak in the next decade or so — and after that, "there's going to be a pretty fast decline on the other side", he says. "That's when there's going to be a rude awakening for the United States." He expects that gas prices will rise steeply, and that the nation may end up building more gas-powered industrial

plants and vehicles than it will be able to afford to run. "The bottom line is, no matter what happens and how it unfolds," he says, "it cannot be good for the US economy."

If forecasting is difficult for the United States, which can draw on data for tens of thousands of shale-gas wells, the uncertainty is much larger in countries with fewer wells. The EIA has commissioned estimates of world shale potential from Advanced Resources International (ARI), a consultancy in Washington DC, which concluded in 2013 that shale formations worldwide are likely to hold a total of 220 trillion cubic metres of recoverable natural gas¹⁰. At current consumption rates — with natural gas supplying one-quarter of global energy — that would provide a 65-year supply. However, the ARI report does not state a range of uncertainty on its estimates, nor how much gas might be economical to extract.

Such figures are "extremely dubious", argues Stevens. "It's sort of people wetting fingers and waving them in the air." He cites ARI's assessments of Poland, which is estimated to have the largest shale-gas resources in Europe. Between 2011 and 2013, the ARI reduced its estimate for Poland's most promising areas by one-third, saying that some test wells had yielded less than anticipated. Meanwhile, the Polish Geological Institute did its own study¹¹, calculating that the same regions held less than one-tenth of the gas in ARI's initial estimate.

If gas supplies in the United States dry up faster than expected — or environmental opposition grows stronger — countries such as Poland will be less likely to have their own shale booms, say experts.

For the moment, however, optimism about shale gas reigns — especially in the United States. And that is what worries some energy experts. "There is a huge amount of uncertainty," says Nehring. "The problem is, people say, 'Just give me a number'. Single numbers, even if they're wrong, are a lot more comforting." ■

Mason Inman is a freelance writer in Oakland, California.

1. Patzek, T. W., Male, F. & Marder, M. *Proc. Natl Acad. Sci. USA* **110**, 19731–19736 (2013).
2. Browning, J. et al. *Oil Gas J.* **111** (8), 62–73 (2013).
3. Browning, J. et al. *Oil Gas J.* **111** (9), 88–95 (2013).
4. Browning, J. et al. *Oil Gas J.* **112** (1), 64–73 (2014).
5. Gülen, G., Browning, J., Ikonnikova, S. & Tinker, S. W. *Energy* **60**, 302–315 (2013).
6. Weijermars, R. *Appl. Energy* **124**, 283–297 (2014).
7. Kaiser, M. J. & Yu, Y. *Oil Gas J.* **112** (3), 62–65 (2014).
8. Hughes, J. D. *Drilling Deeper* (Post Carbon Institute, 2014); available at <http://go.nature.com/o84xwk>.
9. Cook, T. & Van Wagener, D. *Improving Well Productivity Based Modeling with the Incorporation of Geologic Dependencies* (EIA, 2014); available at <http://go.nature.com/dmwsdd>.
10. US Energy Information Administration *Technically Recoverable Shale Oil and Shale Gas Resources* (EIA, 2013); available at <http://go.nature.com/mqkmwx>.
11. *Assessment of Shale Gas and Shale Oil Resources of the Lower Paleozoic Baltic–Podlasie–Lublin Basin in Poland — First Report* (Polish Geological Institute, 2012); available at <http://go.nature.com/lw8fg7>.

COMMENT

MATHEMATICS A cautionary tale of set theory and cold war politics **p.34**



EDUCATION Robotics story airs socio-economic bars to careers in science **p.36**

BIODIVERSITY New pasture plants should be screened for invasiveness risk **p.37**

OBITUARY Herman Eisen, immunology pioneer, remembered **p.38**

HELMUT IGNAT/NATL GEOGRAPHIC ROMANIA



White pelicans feed on the Danube delta in Romania, a global biodiversity hotspot.

Protect the world's deltas

Sea-level rise and river engineering spell disaster, say **Liviu Giosan** and colleagues.

More than a century after Mark Twain argued that the Mississippi River could never be tamed by engineers, dams and dikes constrain the great waterway. Its once muddy waters run clearer¹. Starved of sediment and fragmented by economic development, the Mississippi delta downstream is shrinking by thousands of hectares per year. The rich delta ecosystem and the services it provides — storm protection, nutrient and pollution removal and carbon storage — are being destroyed. Fisheries and the bayou cultural heritage are threatened.

Deltas worldwide share this fate². In Pakistan, one-fifth of the Indus delta plain has been eroded since the river was first dammed in 1932. In China, the northern shore of the Yellow River delta has retreated 300 metres each year for the past 35 years.

Rising seas compound the sediment crisis.

Coastal lowlands less than a metre above sea level will be inundated by the turn of the century. Areas at risk of flooding in deltas will grow by 50%². This global scale of delta drowning has been unprecedented in the past 7,000 years.

Better-preserved deltas such as the Danube, the most extensive wetland in the European Union and a global biodiversity hotspot, show what will be lost. Its labyrinth of channels, lakes, marshes and dunes are home to about 2,000 plant and 5,000 animal species, many of them threatened.

More than 500 million people worldwide live on deltas, many in sprawling megacities such as Shanghai, Dhaka and Bangkok. Conservative estimates value major deltas worldwide at trillions of US dollars in terms of economic revenue and ecosystem services.

In Pakistan, people are leaving the lower

Indus delta because saltwater-soaked soils make agriculture difficult. Nearly ten years after hurricanes Katrina and Rita overwhelmed protective levees in New Orleans, Louisiana, the city's population is still 17% lower than in 2005. In other deltas in the tropical-cyclone belt, such as the Ganges-Brahmaputra in India and Bangladesh, and the Irrawaddy in Myanmar, marsh and mangrove destruction has reduced flood protection and damaged communities and economies. The consequences of delta loss are global: huge losses of ecological services, economic and social crises, and large-scale migrations.

STARVED LANDS

We call on the United Nations to establish an international body of experts to coordinate delta-maintenance initiatives worldwide. We urge governments to accelerate scientific

► research and expand monitoring and forecasting programmes, impact studies and public consultations.

Deltas can exist only where rivers discharge enough sand and mud to overcome the currents that wash them away. But fluvial loads have plummeted³. The Nile and the Indus carry 98% and 94% less mud, respectively, than they did a century ago. Deficits afflict waterways on all scales, from continental systems such as the Mississippi (carrying 69% less sediment than when it was first dammed) and the Danube (60% less) to the smaller rivers such as the Rhône (85% less) in Europe. China's Three Gorges Dam will reduce the Yangtze's load to well below its current 66% reduction from when it was first dammed. The Mekong in southeast Asia will follow this downward trend as dams are built in its basin.

Deltas' naturally high subsidence rates are exacerbated by human activities². For example, the Chao Phraya delta in Thailand is sinking by 5–15 centimetres each year because of intense groundwater use. The Po delta in Italy subsided by 3–5 metres in the twentieth century, mainly as a result of methane extraction. As marshes falter, vegetation dies and halts soil formation, speeding land loss.

Initiatives are trying to save deltas. They include Future Earth and Sustainable Deltas 2015, supported by the International Council of Science, and networks such as the Delta Alliance. But their findings are yet to be implemented. Methods such as rebuilding marshes with dredged sediments, installing coastal protection and diverting river channels to build land are being used on or planned for the Mississippi delta, for example, by the Louisiana Coastal Protection and Restoration Authority. It is too soon to know how effective these efforts will be.

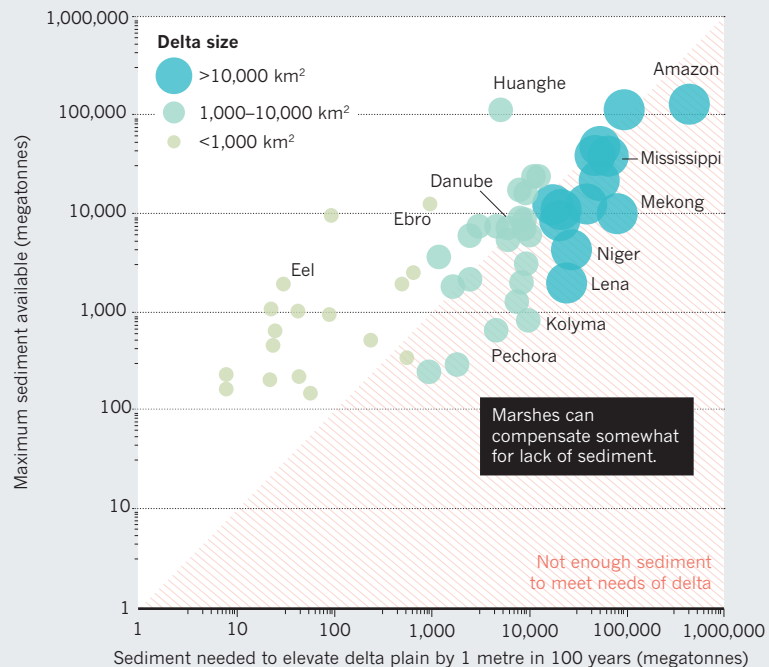
Restoration of drowned delta areas is expensive. The plan for the Mississippi — the most ambitious yet — will run for 50 years, and cost between US\$500 million and \$1.5 billion a year. And it will only stave off future land loss, not recover the vast amount of wetlands already gone.

Knowledge gaps impede the adoption of generic solutions. Delta plains comprise an ever-shifting web of high ground surrounded by low wetlands, shallow lakes and lagoons. Rising sea levels, subsidence and marsh destruction expand lakes. Riverbanks and raised sand ridges become more isolated and vulnerable. Yet there is scarcely any real-time monitoring of water and sediment redistribution, subsidence or ecosystem dynamics. For most deltas, sediment budgets and ratios of mud, sand and organic soils in the upper few metres of a delta plain — crucial for preventing drowning — have yet to be assessed.

The proportion of sediment that ends up on and off shore is poorly quantified^{4–6}. The majority is washed away because delta plains are inefficient sediment traps; even more so

IN THE RED

Most large- and medium-sized deltas cannot grow fast enough to keep up with sea-level rise in the next century. Dams reduce sediment load further and push more deltas into the red.



SOURCES: LG/J/S/SC/J/D.

when they host engineered infrastructure.

Sediment retention depends on particle size, delta structure and vegetation. Sand makes up less than 10% of the load in most rivers. It is important for building land, especially during the early life of a delta. Sandy deltas such as the Krishna and Godavari in India or the Doce and São Francisco in Brazil have more land above sea level (30% by volume on average) than do muddy ones such as the Danube, Ebro and Rhône (10%).

Mud is just as important in building and maintaining large deltas, even though it is more mobile. It improves soil consistency and longevity, keeping marshes and mangroves healthy^{7,8}. Yet little is known about the optimal amount of mud needed to support freshwater and saltwater ecosystems.

The majority of deltas are on course to drown. Even without coastal erosion, marine inundation will exceed the potential for growth in all deltas larger than 10,000 square kilometres and most of those from 1,000 to 10,000 square kilometres (see 'In the red').

The reasons that some deltas grew so large hold lessons for the future. Deltaic 'lobes' may persist when rivers change course. Several channels remain active in the Danube, Niger and Mekong deltas, for instance. Coastal barrier islands protect inland areas from erosion by the sea in the Mississippi and Danube deltas. And sediment loads were boosted by historical deforestation and agricultural practices in Asia, Europe and the United States.

By contrast, deltas less than 1,000 square kilometres are nearly all smaller today than

would have been predicted from their sediment loads pre-damming. They have lost a lot of sediment at the coast and have few lakes and lagoons to trap mud. Although small sandy deltas might survive for a while, muddy ones such as the Ebro or Rhône may quickly lose to the sea. Much of the Po delta is already embanked — below sea level and surrounded by artificial banks, levees and pumps.

The health of wetlands is crucial to delta resilience². Decaying vegetation can compensate for fluvial sediment loss. But we do not know enough about the volume and extent of organic soils, nor about the optimal sediment supply, salinity and nutrients for wetland health.

WORK WITH NATURE

Maintenance strategies can limit delta drowning^{4,5}. These include boosting sediments in river waters, diverting them before they reach the coast, optimizing the trapping abilities of the delta plain and keeping wetlands healthy^{2,4,5–9}. Starting upstream, sluices or flushing mechanisms allow sediment to flow past dams¹⁰. On small rivers, dredging or removing dams may increase sediment flow.

Within the delta, soft engineering strategies should take advantage of natural processes and feedbacks (see 'Fixing the flow'). Multiple channels can be dug to spread sediments across the delta plain⁵ and build marshes, fill ponds and extend lobes⁴. Deliberately breaching levees during floods or redirecting a river to fill basins can enhance natural backwater effects and promote wetland formation.

FIXING THE FLOW

Delta maintenance should mimic natural processes by (1) cutting new channels, (2) breaking banks to build crevasses, (3) constructing small internal deltas in lakes and lagoons or (4) creating new larger lobes in areas protected from waves and tides.

CHANNELIZATION

Used in the Danube delta



CREVASSE SPLAYS

Being tried on the Mississippi delta



INTERNAL SUBDELTA





As seen in the Atchafalaya basin



LOBE BUILDING

Used in the Yellow River delta



 Maintenance solutions  Coastal flatlands  Channels with raised banks  Lakes, lagoons, submarine deltas

The success of some historical practices is encouraging. In the Danube delta, dredging of new fishing channels off the main river has boosted sediment trapping and local livelihoods⁵. On the Ebro delta, the control of river-sediment supply for rice agriculture has increased mud retention on the plain, so that in places the land is rising faster now than the sea level and the rate of subsidence⁹.

At the coast, hard barriers, levees, dikes and locks may be inevitable. Massive projects are expensive — the enclosure of the Rhine–Meuse delta in the Netherlands cost €5 billion (US\$6.3 billion). Where retreat is an option, reshaping the coast as sandy beaches, barriers and mudflats may allow the landscape to adapt to higher sea levels. Urbanized deltas will feel the most acute pressures. In cities such as Shanghai, Bangkok and Dhaka, infrastructure and architecture will have to adapt to recurrent floods and evacuation plans.

The level of protection possible will depend on a region's wealth². Difficult decisions on which lands to preserve and which to abandon are unavoidable. Pakistan cannot afford to revive the Indus delta without jeopardizing its irrigation system, one of the largest in the world. For rivers such as the Mekong, Danube or Indus, which cross multiple countries, negotiations on sediment rights, like those on water allocation, will be complex.

The window to stabilize the world's deltas is closing fast. By 2100, land losses from rising sea levels alone could reach 5% for higher deltas such as the Krishna–Godavari or the Ganges–Brahmaputra, 30% for the Mekong,

Nile and Yellow, and more than 80% for the lower Danube delta. And the sea will keep rising for centuries even if global warming is stemmed and sediment flow to deltas is restored. Restoration schemes must be bigger, faster and better. They must include all deltas, respond to the rapid pace of environmental and economic change, and support data collection, modelling and real-time monitoring.

NEXT STEPS

Plugging the knowledge gaps is the most urgent task. Observations should take advantage of new techniques and sensors. Satellite-based estimates of fluvial discharge or land subsidence are advancing. But more hydrological stations, tidal gauges and subsidence measurements are badly needed.

Biophysical and biogeochemical research on wetland processes must be expanded to address the wide range of deltaic conditions. Agricultural and industrial practices need to be assessed to select sustainable practices.

Governments of delta nations and of countries in their watersheds, helped by international funding agencies, should undertake this enormous effort using standardized methodologies developed under the advice of UN experts. Knowledge and data sharing on a common platform is crucial.

A science-based global strategy for protecting deltas needs to be developed to reduce costs and risks. A UN body of experts should be set up to coordinate existing international and national initiatives. Sustained investment could come from the UN environment

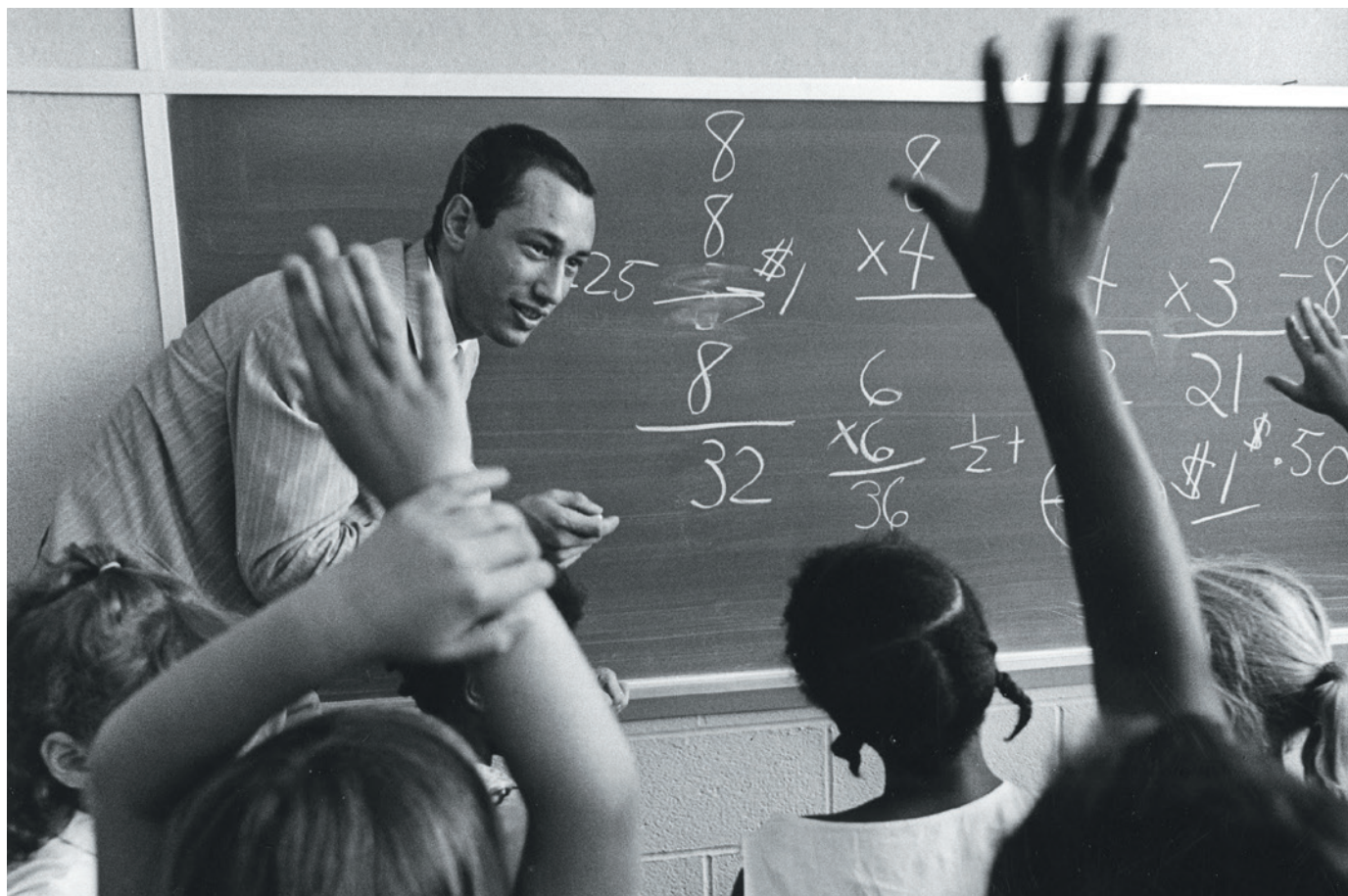
and development programmes (UNEP and UNDP) and the World Bank. Developing economies will need assistance.

River deltas need maintenance now rather than costly restoration later to prevent the collapse of vast expanses of coastline. ■

Liviu Giosan is an associate scientist in geology and geophysics at the Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA. **James Syvitski** is professor of geological sciences at the University of Colorado, Boulder, USA, and chair of the International Geosphere-Biosphere Programme.

Stefan Constantinescu is a lecturer in geography at the University of Bucharest, Romania. **John Day** is emeritus professor of ecology and coastal sciences at Louisiana State University, Baton Rouge, Louisiana, USA. e-mail: lgiosan@whoi.edu

- Day, J., Kemp, G. P., Freeman, A. & Muth, D. P. (eds) *Perspectives on the Restoration of the Mississippi Delta: The Once and Future Delta* (Springer, 2014).
- Syvitski, J. P. M. *et al.* *Nature Geosci.* **2**, 681–686 (2009).
- Syvitski, J. P. M. & Saito, Y. *Glob. Planet. Change* **57**, 261–282 (2007).
- Paola, C. *et al.* *Ann. Rev. Mar. Sci.* **3**, 67–91 (2011).
- Giosan, L., Constantinescu, S., Filip, F. & Deng, B. *Anthropocene* **1**, 35–45 (2013).
- Blum, M. D. & Roberts, H. H. *Nature Geosci.* **2**, 488–491 (2009).
- Day, J. W. Jr *et al.* *Science* **315**, 1679–1684 (2007).
- Kirwan, M. L. & Megonigal, J. P. *Nature* **504**, 53–60 (2013).
- Ibáñez, C., Day, J. W. & Reyes, E. *et al.* *Ecol. Eng.* **65**, 122–130 (2014).
- Kondolf, G. M. *et al.* *Earth's Fut.* **2**, 256–280 (2014).



CINCINNATI MUSEUM CENTER/GETTY

US maths education was kicked into space-age overdrive by the launch of the Soviet Sputnik satellite.

MATHEMATICS

Set theory for six-year-olds

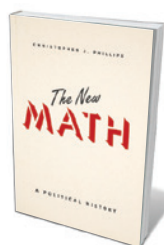
Alex Bellos savours a history of the 'new math' that swept US schools in the 1960s.

It was hailed as a revolutionary mathematics curriculum that prioritized the concept of a set over the concept of a number. 'New math' was introduced to millions of US schoolchildren in the 1960s. Yet in less than a decade it was being lampooned unforgettably by the mathematician and satirist Tom Lehrer. The refrain of his song 'New Math' is: "It's so simple, so very simple, that only a child can do it!"

Teaching set theory to six-year-olds as a way into mathematics, from arithmetic on, would come to be considered misguided, overcomplicating the basics and introducing jargon. Yet however radical a shift it represented, the new math was not a liberal experiment. It was, writes science historian Christopher Phillips in *The New Math*, an essentially conservative cold-war initiative. In 1957, the Soviet Union launched Sputnik, the first Earth-orbiting satellite. The United States was spooked (and

not for the last time) by the risk of falling behind in science, leading to a political consensus that maths teaching needed to be reformed to produce students with sharper cognitive skills.

As Phillips explains, rather than starting a maths class by learning to count, children were asked to think about sets of objects, such as the set of animals on a farm and the set of sweets in a jar. If you paired each animal with a sweet, and there were sweets left over, that set was greater than the set of animals — and vice versa. Perfectly paired-up sets were 'equivalent'. Only now was the concept of number



The New Math: A Political History
CHRISTOPHER J. PHILLIPS
University of Chicago Press: 2014.

introduced, as the property shared by equivalent sets, represented by a symbol called a numeral. The number 1, for example, was precisely that which all sets with a single member had in common.

The New Math charts the method's rise and fall through the story of the School Mathematics Study Group (MSG), an association of mathematicians and teachers formed in early 1958 and funded by the US National Science Foundation. Awash with money, the MSG came up with progressive new ideas for "how mathematics could and should cultivate mental habits"; this, Phillips writes, is "the only reform effort that rooted federal money directly and explicitly in claims about how students should learn to think". The MSG concluded that maths lessons should be less about number crunching and more about abstract thinking. A tension between rote learning and puzzle-solving lies at the heart of maths

pedagogy, but never has practice swung so far from arithmetic drills as during the new-math era. The SMSG's textbook for junior high schools began, for example, with the story of a mathematician on a plane telling his neighbour that counting with numbers is best done by machines and that his job is "logical reasoning".

One reason for the SMSG favouring abstraction over calculation was that it was led by professional mathematicians, not teachers. Many were deeply influenced by Nicolas Bourbaki, the pseudonym for a group of ideological French mathematicians of the 1930s who argued that mathematics was about structure, and that set theory underlay the structure of numbers. 'Modern' maths, argued the SMSG, was needed for space-age science.

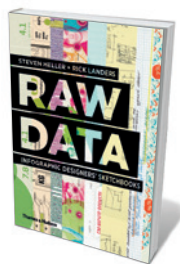
New math was a voluntary programme, and in terms of take-up it was a huge success. By 1965, it was estimated, nearly 75% of US high-school pupils and 40% of elementary-school pupils were studying a version of it; a decade later, it was still taught in an estimated 85% of schools. Major educational publishers sold millions of books influenced by the SMSG's ideas. But the backlash, which had started in the early 1960s, had become fierce by the 1970s. And as Phillips points out, the reasons were as much to do with politics as with pedagogy.

Phillips notes the irony of a nationalistic, cold-war-inspired project introduced to promote rigorous thinking becoming the target of conservatives less than two decades later, partly because they claimed that it failed to impart mental rigour. The 1970s 'back to basics' movement — a decentralized collection of vocal parents horrified at the lack of memorization drills — echoed broader political shifts. People were less inclined to accept diktats from elite authorities such as Washington DC politicians and Ivy League academics. The new math was criticized, possibly unfairly, for declining test scores, rowdy classrooms and children's poor computational skills, and its name became a byword for crazy progressive reforms. The SMSG shut down in 1972, although many schools that had invested in new math had to keep using the same textbooks until the end of the 1970s and beyond. (By then, the method had spread around the world. I was taught set theory at primary school in Scotland in the late 1970s.)

Phillips reminds us in his fascinating book that even though mathematics is supposed to be apolitical, the teaching of it is anything but. ■

Alex Bellos is the author of *Alex Through the Looking-Glass* (US title: *The Grapes of Math*). He is based in London.
e-mail: alexanderbellos@gmail.com

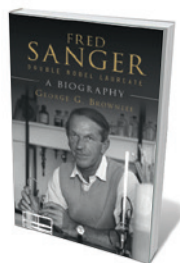
Books in brief



Raw Data: Infographic Designers' Sketchbooks

Steven Heller and Rick Landers THAMES AND HUDSON (2014)

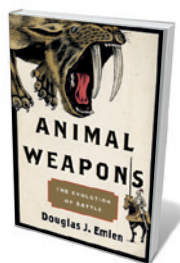
An artist's sketchbook is a peek at cognition's wilder shores: the on-the-fly observations and dogged experimentation that feed the final picture. In this splendid compilation, Steven Heller and Rick Landers riffle through the pages of 73 infographic artists — the wizards who translate numbers into graphics. Juxtaposing doodles with smooth finales, the book explores the work of stars such as Nigel Holmes, whose 'meat and two veg' head reworks the food pyramid in wacky homage to the 'cornucopia' portraits of sixteenth-century artist Giuseppe Arcimboldo.



Fred Sanger — Double Nobel Laureate: A Biography

George G. Brownlee CAMBRIDGE UNIVERSITY PRESS (2014)

The biochemist Frederick Sanger, whose DNA sequencing method made the Human Genome Project possible, died last year. One of only four scientists to win two Nobel prizes, Sanger is revealed in this slim biography by George Brownlee (who studied under him) as inherently modest about his landmark achievements in protein and DNA sequencing. Included are a revealing and fascinating 1992 interview with Sanger that elucidates how he hit on his discoveries, and encomia from the likes of fellow laureates Paul Berg, Elizabeth Blackburn and Paul Nurse.



Animal Weapons: The Evolution of Battle

Douglas J. Emlen HENRY HOLT (2014)

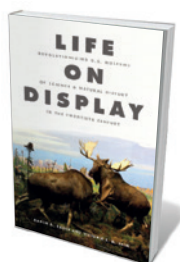
It began with dung beetles. Evolutionary biologist Douglas Emlen's self-confessed fixation on 'extreme' animal weapons was first channelled into research on these horned insects. In this original study, Emlen tours offensive and defensive anatomy and behaviours across aeons and taxa, from *Tyrannosaurus rex*'s fearsome teeth to ibex horns and amphibian poisons. He sharpens the discussion by interweaving parallels with humanity's own evolving arsenal, including weapons of mass destruction — the most extreme of all arms, which if deployed cancel out the very concept of battle.



Banned: A History of Pesticides and the Science of Toxicology

Frederick Rowe Davis YALE UNIVERSITY PRESS (2014)

Historian Frederick Davis takes on the ambitious task of writing the history of pesticides before and in the wake of Rachel Carson's iconic 1962 book *Silent Spring*. In a treatment dense with references, acronyms, chemical names and technical discussions of lethal doses, Davis offers a keen historical perspective on the compounds that big agriculture has used to keep bugs at bay. He forms a scholarly indictment of a system that heeded Carson's superficial message in banning DDT, but ignored her underlying warning that existing systems for managing pesticides are often inadequate.



Life on Display: Revolutionizing U.S. Museums of Science and Natural History in the Twentieth Century

Karen A. Rader and Victoria E. M. Cain UNIVERSITY OF CHICAGO PRESS (2014)

The exquisite dioramas in New York's American Museum of Natural History have wowed crowds since the early twentieth century. But as historians Karen Rader and Victoria Cain reveal in this cogent study, they were part of a broader revolution: the "New Museum Idea", which saw "small machines" and dynamic models supersede dusty cases. The behind-the-scenes struggles between 'edutainers' and serious museum researchers were, they show, no less dynamic. *Barbara Kiser*



The Falcon Robotics team in 2010.

ROBOTICS

Bottom-up innovation

Noel Sharkey ponders a riveting story of how social inequity can trump youth and scientific brilliance.

Joshua Davis's *Spare Parts* is in part a feel-good story of education in action. It tells how four ingenious, but educationally and financially deprived, teenagers were drawn into competitive robotics by exceptional teachers. Pacy and inspirational, it is also a disillusioning tale of how achievement can falter in the face of socio-economic inequities.

The story began in 2004. Cristian Arcega, Lorenzo Santillan, Luis Aranda and Oscar Vazquez were Mexican-born students at Carl Hayden Community High School, at the rough end of Phoenix, Arizona. None had ever seen the sea and not all were viewed as technologically gifted, but science teachers Allan Cameron and Fredi Lajvardi encouraged them to build a robotic submarine. The aim: enter the prestigious, NASA-sponsored Marine Advanced Technology Education Robotics Competition. With a budget of US\$1,000, the team, calling itself Falcon Robotics, set out to compete against the corporate-funded likes of the Massachusetts Institute of Technology in Cambridge.

The four, Davis shows, became a case study in how creativity and imagination can triumph over riches and education. Their infectious enthusiasm got them cheap and free equipment. Stinky, their entry, looked more like plumbing than a robot. Its metre-tall rectangular frame was made from PVC pipes glued together; these provided housing for propellers, sensors, pumps and cameras. Their competitors had well-engineered craft



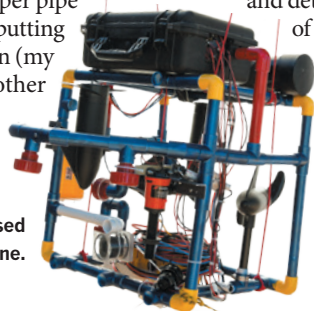
Spare Parts: Four Undocumented Teenagers, One Ugly Robot, and the Battle for the American Dream

JOSHUA DAVIS
Farrar, Straus & Giroux:
2014.

built with aluminium hulls or sophisticated materials. The teams were given a seven-part task centring on a mock-up of a German U-boat. Challenges ranged from measuring the boat's length and navigating inside to recovering 'lost' equipment. The Falcons had had limited time to test Stinky in a diving pool. So they improvised. To calculate depth, they adapted a laser range-finder; to determine length, they hooked a tape measure to the U-boat. A camera on Stinky's underside was used to find objects, and a makeshift robot arm picked them up. The hardest task involved sucking oil out of a barrel and bringing it to the surface. Santillan managed it by gluing a copper pipe to a US\$35 suction pump and putting it inside a balloon. This solution (my favourite) worked — and no other team succeeded.

At the competition in California, the four were stunned

'Stinky', the improvised robotic submarine.



by the other robots. Their hearts sank on hearing that they had won a special award. But NASA judge Lisa Spence loved Stinky's elegant simplicity, and they were pronounced overall winners. At that point in *Spare Parts*, they run down the beach, yelling at the tops of their voices. But this American dream slid into nightmare. Only Aranda was a legal resident. The other three were illegal immigrants, who had come over the border as children. They had no university funding and scant job prospects.

Davis's entry into the story provides some relief. A journalist, he saw a badly written press release about the team that haunted him. Finally he wrote a *Wired* magazine feature that caught public attention. A college fund was collected. Arcega, one of the brightest students in the school district, and team leader Vazquez went to Arizona State University. Santillan and Aranda trained in catering.

But as Davis eloquently reveals, undocumented Mexican Americans have a harrowing time. In Arizona, civilian 'immigration posses' recruited by a local sheriff hunt Latino areas for people committing small offences, and neocon senators push for stricter laws. In 2006, Republican senator Dean Martin pushed through Proposition 300, which stopped state universities offering reduced fees to undocumented residents. This tripled fees for the two at university. Arcega was forced to drop out. Only Vazquez managed to gather enough funding. He married, had a baby, graduated with the highest honours and was commended in front of President Barack Obama and a 70,000-strong audience.

He then returned to Mexico to apply for a US visa. Separated from his wife and child, he picked beans for less than \$4 a day. His application and appeal were refused, and he was banned from entering the United States for ten years. After a flood of public and media support, he was finally granted permanent residency. He is now an engineer. Santillan is a sous chef. The brilliant Arcega lives at home, inventing in his makeshift lab. Aranda works nights, emptying dustbins.

This is perhaps the most gripping popular-science book I have read. Unsurprisingly, the story has been picked up by filmmakers: Sean McNamara's *Spare Parts* will be out in 2015; Mary Mazzio's acclaimed documentary *Underwater Dreams* (go.nature.com/rivaauv) was released in July. The narrative drives itself, but it took Davis's considerable skill and detailed research to pull off a tale

of scientific yearning thwarted by political and social impediments with such pathos. ■

Noel Sharkey is emeritus professor of artificial intelligence and robotics at the University of Sheffield, UK.
e-mail: noel@dc.shef.ac.uk

Correspondence

Conservation: listen to more voices

Heather Tallis and her 239 co-signatories rightly argue that conservation science and practice would benefit from a more inclusive representation of scientists and practitioners (*Nature* **515**, 27–28; 2014). Their call is weakened, however, because it is dominated by voices from the United States (68% of the authors). A further 10% are from Australia.

Non-English-speaking and developing regions are especially poorly represented, with just 5% of the authors coming from continental Europe and 1% from Asia; there are none from China, India, Russia, the Middle East or Japan. This is in no way representative of the global distribution of conservation need or expertise, or of the scientific literature published on conservation topics.

David A. Wardle *Swedish University of Agricultural Sciences, Umeå, Sweden.*
david.wardle@slu.se

Conservation: stop profit trumping all

The barriers to diversifying voices and values in conservation and its governance run deep (H. Tallis *et al.* *Nature* **515**, 27–28; 2014). Anyone who has sat on a diversity and equity committee knows how hard it can be to implement changes in practice.

Barriers include systemic racism and a dominant political-economic system that champions profitability above all. Furthermore, regulations governing issues from land tenure to wilderness acts are relics of colonial administrations — hangovers from elite Euro-American visions of property and land management.

What is needed to make a world where a genuine diversity of views can take root? We suggest that conservation should embrace something

akin to the climate justice movement, in which scientists, activists, academics, farmers, indigenous peoples, urban and rural populations demand not just emissions reductions, but a dramatic redistribution of wealth and power.

Jessica Dempsey *University of Victoria, Canada.*

Rosemary Collard *Concordia University, Montreal, Canada.*

Juanita Sundberg *University of British Columbia, Vancouver, Canada.*

jdempsey@uvic.ca

Conservation: more than inclusivity

Tallis *et al.* missed two opportunities to strengthen their call for greater inclusion in the conservation-biology debate (*Nature* **515**, 27–28; 2014).

First, they could have cited more female authors. Of 37 references, only 9 are first-authored by women (of the 103 authors referenced 28 are women). Thus the authors replicate a form of exclusion they critique. Citing members of under-represented groups as legitimate sources of knowledge is essential to making a discipline inclusive.

Second, the letter did not engage with scholars in the humanities with expertise in values, politics and power — major areas of contention. For example, the authors ignore the fact that Yellowstone National Park's founding, like that of some other national parks, depended on the US government's violent expulsion of Native Americans.

Our characterization of historical and contemporary conservation movements is indeed too Western, too white, too wealthy — too Thoreau, Muir and Leopold, with an occasional Carson thrown in. But when the authors call for “an end to the fighting”, they should not confuse inclusion with harmony. Nor is inclusion itself a

guarantee of equal voice or equal representation.

Laura J. Martin, Sara B.

Pritchard *Cornell University, Ithaca, New York, USA.*

laura.jane.martin@gmail.com

Conservation: focus on implementation

High-impact conservation science may offer insights for policy-makers and funders, but in our experience it is of limited value on the ground. To protect biodiversity, research needs to focus more on how conservation can be implemented (B. Pressey *et al.* *Nature* **515**, 28–31; 2014).

Biodiversity conservation is achieved by activities that abate threats to species and their habitats, and are mostly done by local practitioners. Conservation science helps, in principle, by determining which actions are likely to be effective. This work is mostly done by academics, who are, however, judged by citation rather than conservation impact.

What conservation practitioners need is practical advice rather than academic analysis and assessment (see E. Meijaard and D. Sheil *Biodiversity Conserv.* **16**, 3053–3065; 2007; and M. Cardillo and E. Meijaard *Trends Ecol. Evol.* **27**, 167–171; 2012).

Take our Indonesian-language book on how timber concessions can improve wildlife management (E. Meijaard *et al.* *Hutan pasca pemanenan*; CIFOR/UNESCO, 2006). The book has been used widely in teaching and training, and for informing changes to forestry practices and strategies — including Indonesia's ecosystems restoration initiative. It has been downloaded 35,000 times — yet has garnered only four citations. We need to develop a conservation science that does far more to achieve conservation.

Erik Meijaard *People and Nature Consulting International, Jakarta, and CIFOR, Indonesia.*

Douglas Sheil *Norwegian University of Life Sciences, Ås, Norway, and CIFOR, Indonesia.*
Marcel Cardillo *Australian National University, Canberra, Australia.*
emeijaard@gmail.com

New pasture plants pose weed risk

Many new plant varieties that are sold for livestock pasture pose a weed risk that jeopardizes their purpose — the sustainable intensification of agriculture — by increasing the environmental costs of food production. We urge governments to include potential environmental damage when screening new pasture varieties and to introduce a ‘polluter pays’ penalty system.

More than 90% of new pasture-plant species are invasive weeds with characteristics that include fast growth and wide-ranging environmental tolerance (see D. A. Driscoll *et al.* *Proc. Natl Acad. Sci. USA* **111**, 16622–16627; 2014). The consequences have been disastrous in some countries, including Australia and the United States, where buffel grass, for example, increases fire risk and transforms ecosystems.

Worldwide, limited regulation of new pasture varieties places the environment at increased weed risk. We suggest that agribusinesses should be held financially accountable for environmental damage that their products cause, providing incentives to stem the threat from invasive pasture plants.

Don Driscoll, Jane Catford *Australian National University, Canberra, Australia.*
don.driscoll@anu.edu.au

CONTRIBUTIONS

Correspondence may be submitted to correspondence@nature.com after consulting the author guidelines at <http://go.nature.com/cmchno>.

Herman Eisen

(1918–2014)

Immunologist and educator who discovered fundamentals of antibody binding.

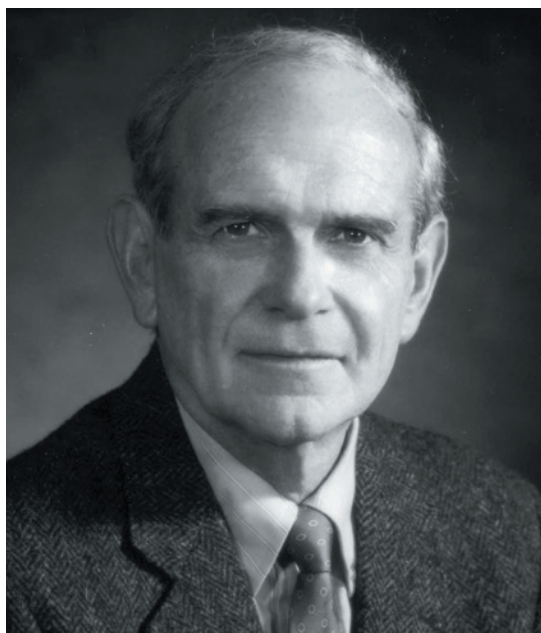
Herman Eisen made fundamental contributions to our understanding of how the body's immune system recognizes foreign structures. In an almost 70-year career, he showed how antibodies bind to antigens (surface molecules on viruses and bacteria) and measured how and why such binding strengthens, a concept that now dominates the field of vaccine design.

Eisen died in Cambridge, Massachusetts, on 2 November. He was born in 1918 into the Jewish community of Brooklyn, New York, one of four children, to parents who had emigrated from Eastern Europe. "I grew up with a sense that anti-Semitism in the world around us was pervasive, regarded as a fact of life, like birth and death," he later wrote. At age 16, he was interested in chemistry but enrolled as a premedical student at New York University (NYU), persuaded by his father that "large companies in the chemicals industry did not hire Jews. In medicine, however, a somewhat related field, one's destiny was in one's own hands."

While recuperating from tuberculosis during his university years, Eisen was deeply affected by two works: Sinclair Lewis's *Arrowsmith* (1925) and Charles Darwin's *On the Origin of Species* (1859). He was inspired by a seminar in which he learnt that moving a hydroxyl group on a benzene substituent from one carbon atom to another dramatically changed the compound's effect on a cat's blood pressure. It revealed to him the connection between molecular structure and biological function. As a resident in pathology at Columbia University in New York from 1944 to 1946, Eisen was introduced to immunology in the laboratory of Michael Heidelberger, whose work had enabled the first demonstration that antibodies were proteins.

During his time as a medical resident at Bellevue Hospital Center, NYU's main teaching hospital, Eisen worked as a ship's surgeon on a voyage to South America. Between ministering to crewmen with minor injuries or syphilis, Eisen read Karl Landsteiner's superb 1936 summary of what was then known about immunology, *The Specificity of Serological Reactions*. The monograph argued for molecular complementarity as the basis for antigen recognition, a concept that inspired much of Herman's later work.

In 1948, Eisen received a senior National Institutes of Health fellowship grant of



US\$3,600 per year, freeing him to do full-time research and to marry Natalie Aronson, a paediatrician, and start what became a family of five children. Working in the biochemistry department at NYU, Eisen, with Fred Karush, showed that antibodies typically possess two active sites that can bind antigens. This permits the formation of aggregates that are eliminated from the body. However, if such complexes lodge in joints or in the kidney, they can cause disease.

Next, Eisen moved to NYU's department of industrial medicine to explore allergic skin reactions to dinitrobenzenes. Working on his own skin and that of guinea pigs, he discovered that only those chemicals that are capable of forming covalent bonds to skin proteins cause the characteristic itchy rash. These studies led to his appointment in 1955 as professor of dermatology at Washington University in St. Louis, Missouri; he became chair of microbiology there in 1961.

Eisen was one of the handful of like-minded pioneers intent on understanding immune recognition who met periodically in a series of 'antibody workshops'. These advanced the problem of antibody diversity — how the body raises defences against different invaders — much as the RNA Tie Club, founded by James Watson and George Gamow, coalesced the pioneers of molecular biology.

At Washington University, Eisen developed

an ultrasensitive method to measure the strength of binding — the binding constant, or affinity — of antibodies for chemical entities such as dinitrophenol. He showed that binding becomes stronger the more time has elapsed since the first exposure to an antigen, because cells producing more tightly binding antibodies are preferentially stimulated as the amount of antigen diminishes. After repeated exposure to the same antigen ('booster' immunization) the tightly binding antibodies are synthesized without delay.

In 1973, he was recruited as a founding member to the newly established Center for Cancer Research (now the Koch Institute for Integrative Cancer Research) at the Massachusetts Institute of Technology in Cambridge. His research interests broadened to include recognition by receptors for antigens on blood cells known as T lymphocytes. He contributed to the discovery and characterization of the elusive T-cell receptor, and provided experimental and conceptual tools to show the extraordinary sensitivity of 'killer' T lymphocytes, capable of responding to the presence of even a single antigen molecule as a means of eradicating a virus-infected cell.

At his house near Woods Hole, Massachusetts, Eisen indulged his passion for gardening, enjoying the biology and the aesthetics of what he nurtured. The author of an influential textbook, *General Immunology* (1990), Eisen was a sought-after mentor for students, postdoctoral fellows and sabbatical visitors.

With his unwavering curiosity and commitment to research, he was always available to discuss science, scheduling lunch dates to catch up with research areas in which he was interested but not necessarily active. Herman Eisen remained at the frontier of immunology: he was working on a manuscript and e-mailing coworkers the day that he collapsed en route to the gym, aged 96. ■

Lisa Steiner is professor of biology at the Massachusetts Institute of Technology (MIT) in Cambridge, Massachusetts, USA. She worked under Herman Eisen as a postdoctoral researcher between 1962 and 1965, and they were friends for more than 50 years. **Hidde Ploegh** is professor of biology at MIT, and worked with Eisen at the MIT Center for Cancer Research from 1992. emails: lsteiner@mit.edu; ploegh@wi.mit.edu

COURTESY OF MIT

Autophagy transcribed

Two studies find that an intracellular quality-control mechanism called autophagy is regulated by nuclear receptor proteins that govern the expression of autophagy genes. [SEE LETTERS P.108 & P.112](#)

CARMINE SETTEMBRE & ANDREA BALLABIO

During periods of starvation, cells survive by changing their status, minimizing the energy spent on biosynthesis and instead activating catabolic pathways that release energy from intracellular stores. This adaptation requires autophagy¹, a process by which intracellular substrates are transported to a subcellular compartment called the lysosome, then degraded and recycled. Although the mechanisms underlying the regulation of autophagy have long been considered to be post-transcriptional, recent studies² have uncovered a role for transcriptional networks in the process. Two papers^{3,4} in this issue describe a previously unknown transcriptional mechanism that regulates autophagy in response to nutrient availability.

The emerging picture of how autophagy is regulated is of a biphasic mechanism of short- and long-term responses. In the cytoplasm, post-translational protein modifications and protein-protein interactions mediate the rapid induction of the pathway, but nuclear transcriptional mechanisms are necessary for a sustained response². Such transcriptional regulation ensures that the proteins required for lysosome formation and autophagy are produced in appropriate quantities during long periods of nutrient shortage.

FXR is a nuclear receptor protein that is active during normal feeding conditions and regulates bile-acid, lipid and glucose metabolism⁵. Using mouse models and pharmacological approaches, Seok *et al.*³ (page 108) and Lee *et al.*⁴ (page 112) both demonstrated that FXR is a repressor of autophagy in the liver, the first direct evidence of a link between nuclear receptors and autophagy. However, the two studies report different underlying mechanisms for this repression. Seok and colleagues found that active FXR blocks autophagy by inhibiting the transcriptional activity of CREB, a protein that promotes the expression of several autophagy genes. FXR mediates such inhibition by disrupting the functional interaction between CREB and its coactivator protein CRTC2. By contrast, Lee and co-workers showed that FXR binds directly to the promoter-DNA

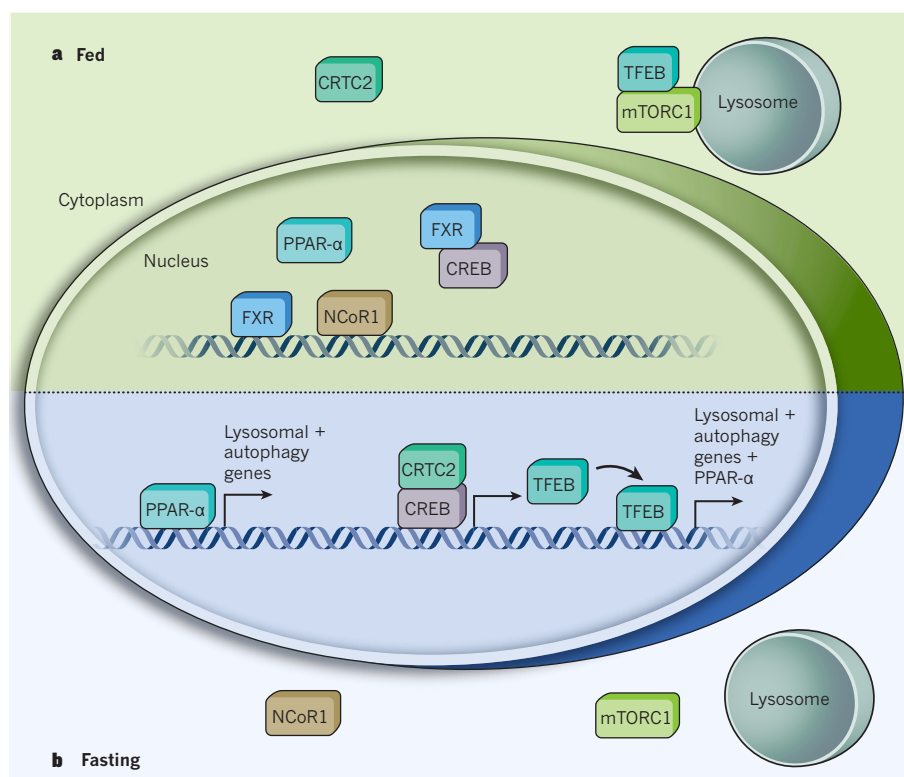


Figure 1 | Regulation of starvation-induced autophagy. Seok *et al.*³ and Lee *et al.*⁴ report that the nuclear receptor protein FXR regulates autophagy. **a**, In normal feeding conditions, FXR inhibits the formation of the CRTC2–CREB protein complex by binding CREB, and competes with another nuclear receptor protein, PPAR- α , for binding sites on gene-promoter regions. Fed conditions also mean that the autophagy-promoting transcription factor TFEB is unable to enter the nucleus because it is bound by the protein mTORC1, which is located on structures called lysosomes. Finally, the transcriptional repressor protein NCoR1 (which is activated by mTORC1) binds DNA. Together, these actions repress the transcription of autophagy genes. **b**, During fasting, FXR is inactive, leading to PPAR- α –promoter binding. This activates the transcription of lysosomal and autophagy genes. In the absence of FXR, the CRTC2–CREB complex forms, which activates transcription of the gene encoding TFEB. This protein is not held by mTORC1 in fasting conditions and can therefore promote transcription of autophagy genes, including that encoding PPAR- α . The co-repressor NCoR1, however, is no longer activated by mTORC1 and is held in the cytoplasm.

regions that regulate the expression of several autophagy genes, leading to their repression.

Notably, Lee *et al.* showed that the binding of FXR to promoter DNA occurs at regions called DR1 sites, which can also be bound by the nuclear receptor protein PPAR- α . Like FXR, PPAR- α is involved in lipid metabolism, but unlike FXR, it is activated by fasting and promotes the production of energy from the degradation of liver fatty acid⁵. The authors found that PPAR- α induces the expression of

autophagy genes, whereas FXR represses it. Thus, the two factors antagonistically regulate the autophagic response to nutrient availability by competing for the same binding sites on DNA.

The apparent discrepancy between the mechanisms proposed in the two papers probably reflects the complexity of the machinery involved in the regulation of autophagy in response to nutrient levels. FXR may repress transcription through

many mechanisms to fine-tune autophagy. Moreover, only a fraction of the DNA sites bound by FXR are bound by CREB, supporting the possibility that, under normal feeding conditions, FXR can inhibit autophagy through CREB-independent mechanisms. It is also possible that the two mechanisms reported are differentially activated in response to different nutritional cues.

The number of autophagy genes whose expression seems to be controlled either directly or indirectly by FXR/CREB and PPAR- α is remarkable, and includes genes involved in several steps of autophagy. For example, Seok *et al.* showed that CREB induces autophagy through the direct transcriptional activation of the TFEB protein, a master regulator of lysosome biogenesis and of autophagy⁶. TFEB also promotes the expression and activity of PPAR- α , thereby activating lipid-degradation pathways in the liver⁷. Thus, these findings suggest that CREB and TFEB, as well as the nuclear receptor proteins FXR and PPAR- α , all belong to the same transcriptional network, which is regulated by nutrients and controls autophagy (Fig. 1). This transcriptional program operates in the nucleus, but has intimate connections with the cytoplasmic pathways that regulate the rapid induction of autophagy. The nutrient-sensing kinase enzyme mTORC1, a regulator of starvation-induced autophagy, mediates these interactions by holding TFEB in the cytoplasm in fed conditions, but permitting nuclear entry during fasting⁸, and regulating nuclear entry of NCoR1, a transcriptional repressor of PPAR- α , in the opposite manner⁹.

A role for autophagy in lipid degradation has already been described — in a process called lipophagy, lipid droplets are internalized in autophagic vesicles and then delivered to lysosomes for degradation¹⁰. The various lipid-degradation mechanisms that operate in different cellular compartments, such as the nucleus, lysosomes, autophagic vesicles and mitochondria, require an integrated regulatory network that is just beginning to emerge. The current studies clearly show that FXR and PPAR- α regulate lipophagy at the transcriptional level.

A link between autophagy and nuclear receptors expands our knowledge of the autophagy repertoire, because nuclear receptors are involved in a multitude of pathways. More studies are needed to fully understand how different physiological and pathological conditions regulate the process. Furthermore, it remains to be determined whether or not the transcriptional mechanisms that mediate the response to starvation are also used to respond to other energy-demanding conditions, such as physical exercise and low temperature, or to diseases such as cancer. Finally, pharmacological modulation of these pathways might point to possible therapeutic strategies for combating a broad range of diseases. ■

Carmine Settembre and Andrea Ballabio are at the Telethon Institute of Genetics and Medicine, Naples 80078, Italy; in the Department of Translational Medicine, Federico II University, Naples; and in the Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA.

1. Rabinowitz, J. D. & White, E. *Science* **330**, 1344–1348 (2010).
2. Füllgrabe, J., Klionsky, D. J. & Joseph, B. *Nature Rev. Mol. Cell Biol.* **15**, 65–74 (2014).

3. Seok, S. *et al. Nature* **516**, 108–111 (2014).
4. Lee, J. M. *et al. Nature* **516**, 112–115 (2014).
5. Evans, R. M. & Mangelsdorf, D. J. *Cell* **157**, 255–266 (2014).
6. Settembre, C. *et al. Science* **332**, 1429–1433 (2011).
7. Settembre, C. *et al. Nature Cell Biol.* **15**, 647–658 (2013).
8. Settembre, C. *et al. EMBO J.* **31**, 1095–1108 (2012).
9. Sengupta, S., Peterson, T. R., Laplante, M., Oh, S. & Sabatini, D. M. *Nature* **468**, 1100–1104 (2010).
10. Singh, R. *et al. Nature* **458**, 1131–1135 (2009).

This article was published online on 12 November 2014.

EARTH SCIENCE

Controls on isotopic gradients in rain

Concentrations of heavy isotopes of hydrogen and oxygen decrease in rain as storms cross land. A model examines the transport of water vapour that causes this effect, and provides insight into past and present climates.

KATHERINE H. FREEMAN

Water vapour is lifted from the sea by evaporation, transported by storms and then released across land as precipitation. As rain falls from clouds, it preferentially removes water containing heavy isotopes, so that both clouds and rain become progressively depleted in deuterium and oxygen-18 as they move across continents. The extent of this ‘rainout’ dominates all the observed trends of continental isotope gradients, but the gradients can vary by up to 50-fold. These variations reflect complex factors that have long eluded simple

quantification. Writing in *Earth and Planetary Science Letters*, Winnick *et al.*¹ address this problem with their report of an improved one-dimensional model that follows rain isotopes along a storm track.

The water-cycle processes that cause isotope ‘sorting’ in rain are well understood, and the resulting isotope gradients in continental precipitation have been extensively documented². At the mid-latitudes, where air temperatures are unstable, atmospheric mixing is particularly pronounced. This lowers isotope gradients as a result of the diffusion-like smoothing of turbulent mixing, a process called eddy diffusion³. And in tropical lowlands such as



Figure 1 | Rain over the Denali National Park, Alaska. Winnick *et al.*¹ have modelled water-vapour transport along the track of a storm.

the Amazon Basin, small humidity gradients across long distances diminish isotope sensitivity to mixing and transport^{4,5}.

Moisture fluxes from plants and evaporated surface water also affect isotope gradients in precipitation. Transpiration — the evaporation of plant water — occurs through leaf stomata, which act like the tips of small capillaries. The narrow plumbing in leaves prevents back-mixing of evaporated water, and so the water vapour that exits from plants has the same isotopic composition as water taken up by their roots. The flux of moisture returned from plants therefore has a modest influence on downstream rainfall isotope gradients, but it does recharge cloud water-vapour pools, and this reduces the extent of rainout⁶. By contrast, evaporation from the surfaces of lakes and rivers strongly sorts water isotopes, and delivers water vapour depleted in heavy isotopes back to clouds^{7,8}. The different amounts and mass effects of surface- and plant-water fluxes in different regions cause continental isotope gradients to be sensitive to the geography of vegetation and aridity.

Winnick *et al.*¹ use their model of water-vapour transport to investigate the dominant factors governing isotope gradients in precipitation. By scaling calculations to the extent of rainout, which they represent using humidity gradients, the authors find that isotopic-gradient variability reflects two things: differing amounts of water vapour returned to clouds (from plants and by evaporation from surface waters, together known as evapotranspiration); and the amount of advection (water-vapour transport by the bulk motion of air) and turbulent mixing in the atmosphere. They conclude that continental isotope gradients correspond to global patterns of air circulation and hydrology, yet are most sensitive to evapotranspiration, transport and mixing in areas near mountains.

Water-isotope records are helpful tools for reconstructing mountain elevation and climate patterns of the past. The isotope patterns of precipitation-derived water that is chemically captured as carbonates in mountain lakes and soils^{9,10}, as well as in preserved plant waxes¹¹, offer geologists unique insight into how fast and how high mountains rose, and help to constrain tectonic models of mountain building. Hydrologic isotope records also capture the impacts of climate change on the water cycle¹².

But water drawn from different sources along shifting storm tracks can confound isotopic records of elevation and climate processes at any one location (see ref. 11, for example). Winnick *et al.* point out that isotope gradients based on data from several locations can help to eliminate source effects and thereby improve reconstructions of past environments. By normalizing data to the spatial scale of humidity gradients, the authors have provided researchers studying palaeoaltimetry and palaeoclimate with a potent quantitative tool for unpacking the interlinked influences of vegetation, climate

and topography on continental hydrology.

Mountain topography (Fig. 1) causes sharp gradients in both the flux and isotopic character of precipitation. Winnick and co-workers show that these gradients are highly sensitive to factors other than elevation, sounding a cautionary note for palaeoaltimetry applications. But they also find that mountainous localities may capture details of climate impacts that are probably missed in less-sensitive lowland regions — which is good news for climate scientists. Nevertheless, climates and mountain waters are not simple systems. Changing temperatures can shift not only the phase of precipitation (whether it snows or rains, for example), but also the extent of isotope sorting during condensation and evaporation. For example, hot climates reduce isotope gradients in water simply because isotope fractionation is less pronounced at higher temperatures.

Of necessity, Winnick and colleagues' model requires many pre-set parameters, including temperature-linked isotope-mass sensitivities, humidity and the ratio of evaporation to transpiration. But as we venture back through the pageantry of past ecology and climates, parameters based on modern data become increasingly inadequate, because life and its habitats have evolved considerably over long periods of time. Furthermore, short-term variations in vegetation, droughts and temperature — which pose major challenges for society as our planet warms — also occurred throughout Earth's history. The model's requirements

therefore bring into focus the need to support isotope records with palaeoclimate and palaeobotanical data. These will help mountains and their isotope gradients to teach us crucial lessons from the past about water, climate and carbon, and how they are linked across continents and the globe. ■

Katherine H. Freeman is in the Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania 16803, USA.
e-mail: khf4@psu.edu

1. Winnick, M. J., Chamberlain, C. P., Caves, J. K. & Welker, J. M. *Earth Planet. Sci. Lett.* **406**, 123–133 (2014).
2. Dansgaard, W. *Tellus* **16**, 436–468 (1964).
3. Eriksson, E. *Tellus* **17**, 498–512 (1965).
4. Gat, J. R. & Matsui, T. *J. Geophys. Res.* **96**, 13179–13188 (1991).
5. Rozanski, K., Araguás-Araguás, L. & Gonfiantini, R. in *Climate Change in Continental Isotopic Records* (eds Swart, P. K., Lohmann, K. C., McKenzie, J. & Savin, S.) 1–36; *Geophys. Monogr.* **78** (Am. Geophys. Union, 1993).
6. Ingraham, N. L. & Taylor, B. E. *Wat. Resour. Res.* **27**, 77–90 (1991).
7. Horita, J. & Wesolowski, D. J. *Geochim. Cosmochim. Acta* **58**, 3425–3437 (1994).
8. Gat, J. R. *Annu. Rev. Earth Planet. Sci.* **24**, 225–262 (1996).
9. Garzione, C. N., Quade, J., DeCelles, P. G. & English, N. B. *Earth Planet. Sci. Lett.* **183**, 215–229 (2000).
10. Rowley, D. B., Pierrehumbert, R. T. & Currie, B. S. *Earth Planet. Sci. Lett.* **188**, 253–268 (2001).
11. Polissar, P. J., Freeman, K. H., Rowley, D. B., McInerney, F. A. & Currie, B. S. *Earth Planet. Sci. Lett.* **287**, 64–76 (2009).
12. Fricke, H. C. & Wing, S. L. *Am. J. Sci.* **304**, 612–635 (2004).

MICROBIOLOGY

A backup for bacteria

The finding that intestinal viruses can substitute for intestinal bacteria to promote the health of their mammalian hosts raises the possibility that viruses in the gut may be beneficial in some circumstances. [SEE LETTER P.94](#)

YAO WANG & JULIE K. PFEIFFER

A dense population of microorganisms inhabits the intestinal tract, including bacteria, archaea, fungi and viruses. Intestinal bacteria benefit the host by aiding nutrition, promoting immune-cell development and protecting from intestinal damage, but it is unclear whether other members of the microbiota have similar roles. Viruses are relatively poorly studied members of the intestinal microbial community, and mammalian intestinal viruses are generally thought to be harmful or neutral to the host. But on page 94 of this issue, Kernbauer *et al.*¹ show that, in the absence of bacteria, mammalian intestinal viruses promote gut homeostasis and protect the intestine from injury and from pathogenic bacteria. Their work raises the

possibility that mammalian intestinal viruses may benefit the host in some instances.

The collection of viruses in the intestine — the intestinal virome — includes viruses that infect bacteria (bacteriophages), archaeal viruses, plant viruses and mammalian viruses^{2–4}. Although bacteriophages are abundant and ever-present inhabitants of the intestine, mammalian viruses are detectable only some of the time. Members of the mammalian virome include viral pathogens that persist after a disease has ended, non-pathogenic viruses that are often detected in healthy individuals, and uncharacterized viruses whose similarity to known viruses in terms of nucleic-acid sequence is low^{5,6}. Murine norovirus (MNV) is a relatively common member of the mouse intestinal virome and several MNV strains have been discovered

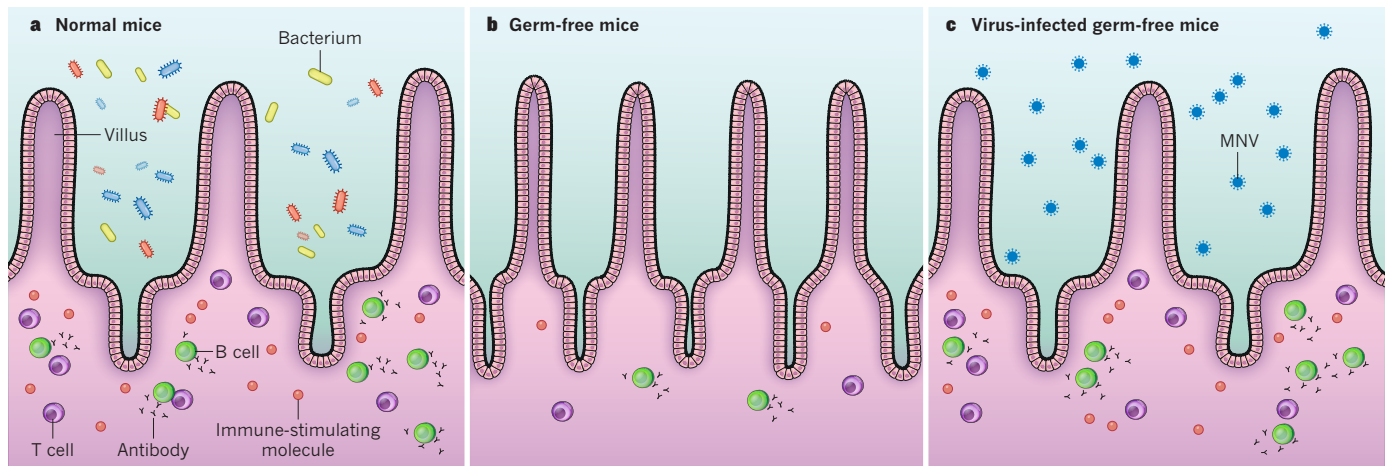


Figure 1 | Beneficial viruses in the mammalian intestine. **a**, The intestine of normal mice harbours a diverse microbial population, including commensal bacteria that maintain tissue architecture and promote the development of lymphocytes (T and B cells), which produce antibodies and other immune-stimulatory molecules, thereby promoting host resistance to injury and to pathogenic bacteria. **b**, Germ-free mice lack microbiota

and have altered intestinal architecture, such as thinner villi, and impaired lymphocyte development. They are thus more susceptible to injury and bacterial pathogens. A similar situation arises in mice treated with antibiotics. **c**, Kernbauer *et al.*¹ show that infection of germ-free or antibiotic-treated mice with murine norovirus (MNV) restores the correct morphology and immunological functions of the intestine.

in animal-research facilities. MNV is generally non-pathogenic in mice that have a functioning immune system, but can cause disease in some immune-deficient mice. Kernbauer and colleagues used three MNV strains as representative members of the mouse intestinal virome to determine whether viruses can influence intestinal homeostasis in 'germ-free' mice that lack a microbiota.

Germ-free mice are microbiologically sterile, and because of the absence of beneficial bacteria in their intestines they have abnormal intestinal morphology and defective development of immune cells called lymphocytes⁷. Antibiotic-treated mice also have several of these abnormalities. The authors found that MNV infection could repair many of the morphological and immunological defects in these mice (Fig. 1). For example, it restored the thickness of the villi (projections of the mucous membrane that lines the intestine) and the number of granules in the Paneth cells of the intestinal lining, which are filled with antimicrobial compounds. Also restored were the numbers of CD4⁺ and CD8⁺ T cells (two classes of lymphocyte), the production of immune-stimulatory molecules by these cells, more-normal levels of antibodies, and a balanced population of innate lymphoid cells, another class of intestinal immune cell. Thus, the presence of a single type of virus reversed many of the defects caused by the lack of bacteria in germ-free or antibiotic-treated mice.

MNV-induced restoration of intestinal cell numbers, cell function and tissue architecture in germ-free mice suggests that the virus can replace some functions of the microbiota. But are there major functional consequences of these MNV-associated changes? For example, can viral infection of germ-free mice limit disease and aid host survival in the face of injury or

infection with pathogenic bacteria? To test this, Kernbauer and colleagues treated normal and antibiotic-treated mice with an intestine-damaging chemical and monitored their survival in the presence or absence of MNV infection. The intestinal injury killed more antibiotic-treated mice than conventional mice, but MNV infection enhanced the survival of antibiotic-treated mice. Similarly, infection with pathogenic bacteria was more damaging in antibiotic-treated mice than in conventional mice, but viral infection reduced disease in antibiotic-treated mice. These data demonstrate that MNV infection can functionally substitute for the microbiota to enhance the animals' health and survival.

This work raises many interesting questions. Mammalian intestinal viruses are generally thought to be detrimental to the host, but previous studies have shown beneficial effects of viral infection in other circumstances. For example, herpesviruses can suppress bacterial infections, and retroviruses were involved in the evolution of the placenta^{8,9}. This work provides the first evidence for beneficial effects of a mammalian intestinal virus, but other intestinal viruses may exist in commensal associations, in which the virus can replicate without harming the host, or in mutualist associations, in which both host and virus benefit. Further investigations will be needed to determine whether such viruses are frequent inhabitants or transient members of the microbiota, and which mammals harbour these viruses and to what effect.

Another outstanding question is, do mammalian intestinal viruses benefit the host in the context of the 'normal' microbiota? The current study examined the beneficial effects of MNV using germ-free mice or mice treated with multiple antibiotics. Presumably, the beneficial effects of a complete microbiota would mask the beneficial effects of a viral infection. However,

the presence of a single type of bacterium, the segmented filamentous bacteria, in the microbiota is known to influence pro-inflammatory responses¹⁰, and particular intestinal viruses might induce similarly specific responses even in a healthy host with a normal microbiota. Indeed, Kernbauer *et al.* found that the three closely related MNV strains induced slightly different responses in the germ-free mice.

Finally, could mammalian intestinal viruses be useful as probiotics? Intestinal inflammation in humans can be caused by antibiotic use, genetic background and other factors. In some cases, such disease can be reversed by faecal transplant — and the consequent transfer of the intestinal microbiota — from a healthy donor. Perhaps certain mammalian intestinal viruses could also ameliorate disease. Moreover, it is possible that some of the observed benefits following faecal transplants are due to viruses present in the donor samples. Future work exploring the intestinal virome should provide some answers to these questions. ■

Yao Wang and Julie K. Pfeiffer are in the Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9048, USA.
e-mail: julie.pfeiffer@utsouthwestern.edu

1. Kernbauer, E., Ding, Y. & Cadwell, K. *Nature* **516**, 94–98 (2014).
2. Reyes, A. *et al.* *Nature* **466**, 334–338 (2010).
3. Handley, S. A. *et al.* *Cell* **151**, 253–266 (2012).
4. Phan, T. G. *et al.* *PLoS Pathog.* **7**, e1002218 (2011).
5. Virgin, H. W. *Cell* **157**, 142–150 (2014).
6. Duerkop, B. A. & Hooper, L. V. *Nature Immunol.* **14**, 654–659 (2013).
7. Hooper, L. V., Littman, D. R. & Macpherson, A. J. *Science* **336**, 1268–1273 (2012).
8. Roossinck, M. J. *Nature Rev. Microbiol.* **9**, 99–108 (2011).
9. Barton, E. S. *et al.* *Nature* **447**, 326–329 (2007).
10. Ivanov, I. I. *et al.* *Cell* **139**, 485–498 (2009).

This article was published online on 19 November 2014.

ASTROPHYSICS

Stars fight back

Galaxies contain fewer stars than predicted. The discovery of a massive galactic outflow of molecular gas in a compact galaxy, which forms stars 100 times faster than the Milky Way, may help to explain why. [SEE LETTER P.68](#)

PHILIP F. HOPKINS

The past two decades have seen a revolution in our understanding of the origins and evolution of the cosmos. The standard hypothesis of a Universe dominated by 'cold dark matter' has been extraordinarily successful in explaining a vast number of astronomical observations. But unsolved puzzles abound when we consider what cosmologists call small scales: namely, single galaxies. According to the simplest cosmological models, nearly all the baryonic matter — the 'normal' matter in the Universe — should sit at the centres of galaxies in the form of stars. But baryonic matter apparently does no such thing. On page 68 of this issue, Geach *et al.*¹ describe the discovery of a galaxy caught in the act of violently expelling its baryonic matter, revealing a process that may be crucial to solving this mystery.

The problem is an old one. Regardless of the nature of dark matter, many diverse observations indicate that only a few per cent of the normal matter in the Universe is in the form of stars and gas concentrated inside galaxies. If we start from the simplest assumptions, with

a certain amount of gas in the early Universe, and evolve it forward in time according to the laws of gravity and chemistry, the gas will eventually condense into increasingly dense objects, forming galaxies and, within those galaxies, stars. Inevitably, this leads to the prediction that nearly all baryonic matter should be in stars. However, only part of it is. In galaxies such as the Milky Way, in which the conversion of gas into stars is known to be relatively efficient, only 10–20% of the normal matter that gravity should have dragged in has actually been retained inside the galaxy (mostly in stars). The rest is thought to be outside the galaxy, in the diffuse intergalactic medium.

How might this occur? Theorists have long speculated that feedback mechanisms may be important. Stars are not simply passive spectators. When massive stars form, they soon explode as supernovae, injecting tremendous energy into the surrounding gas. Because this energy is much larger than the gravitational energy holding that gas in place, it is clear that small galaxies should be able to eject most of their normal matter in a galactic-scale outflow after forming just a small number of stars². However, in more-massive galaxies

(those bigger than our own), this energy source seems to be insufficient for such gas ejection. One idea is that the feedback in these systems may come from the energy released as matter falls (accretes) into the supermassive black holes — with masses from one million to ten trillion times that of the Sun — that are known to lie at the centres of these galaxies^{3,4}. These accreting black holes shine as quasars or their less luminous analogues, active galactic nuclei (AGNs), and are among the most energetic objects in the Universe.

Geach and colleagues' team has been working to uncover the effects of feedback on galaxies. But it is a challenging task. Although many observations have shown that galactic outflows (Fig. 1) are ubiquitous when galaxies are rapidly forming stars⁵, it has been difficult to measure the mass of these winds, and hence the momentum and energy required to form them. The observations have largely indicated the signatures of galactic outflows in spectroscopic features linked to the wind's absorption of galactic light, because the wind lies between the galaxy and the observers. But turning those results into a mass requires detailed knowledge of where exactly the wind lies, and what fraction of the wind is in the precise atomic- and molecular-gas species and ionization state associated with the spectroscopic features.

With their new observations of a galactic outflow, Geach *et al.* have obtained a much more direct tracer of the mass of an outflow than has been possible so far. By measuring the emission luminosity of carbon monoxide, which is known to be an accurate tracer of the total mass of dense molecular gas under interstellar conditions, the authors realized that most of the material in this outflow was 'hiding' in the cold molecular phase, allowing several properties of the outflow to be determined. Measuring the outflow rate of any galaxy, especially one so far away (the light from the galaxy reported here was emitted when the Universe was approximately half its current age) would be a great achievement, but what the authors found was much more remarkable.

The galaxy, dubbed SDSS J0905 + 57, drives a wind with an incredible velocity. The speed of the material flying out from the galaxy reaches 2,500 kilometres per second, much greater than the speeds of about 200–500 km s⁻¹ typically seen around more mundane galaxies. The outflow rate is roughly 100 solar masses per year, the same order of magnitude as the rate at which the galaxy is forming stars (and 100 times higher than the rate of star formation in the Milky Way). If the galaxy can maintain this prodigious outflow, it could empty its reservoir of molecular gas in 10 million years. The wind's velocity places it among the most extreme known.

How could such an outflow be powered? The authors see no AGN in the system, and the energy in supernovae seems insufficient to drive the outflow. But another remarkable



Figure 1 | A galactic outflow. This image of starburst galaxy M82 shows its disk of gas and stars (blue) and a perpendicular outflow of ionized gas (red).

J. GALLAGER (UNIV. WIS.), M. MOUNTAIN (STSC), P. PUXLEY (NSF)/HUBBLE HERITAGE TEAM (STSC/AURA/ESA/NASA)

NEUROBIOLOGY

property of the galaxy may suggest an answer. SDSS J0905 + 57 is extremely compact: half of its huge star-formation rate comes from a region less than 100 parsecs across (the Milky Way is about 20,000 parsecs across). With so much light being emitted from such a small region, the momentum being imparted by the starlight itself on any nearby gas becomes a highly powerful force. The authors show that this 'radiation pressure' could indeed accelerate the wind to such a high velocity that it would sweep up most of the molecular gas. This would also represent a gentler way of pushing on the gas than through supernovae, which generate strong shocks. Such a mechanism would avoid destroying the molecular bonds in the gas. A similar process leads to the Eddington limit — the maximum luminosity of a star. What the authors see here is a truly 'Eddington-limited galaxy'.

Such a radiation-pressure mechanism represents an exciting and unusual channel for feedback. Although radiation pressure has been proposed as a potential feedback channel⁶, it has only just begun to be included in models, and has remained observationally elusive. That may change rapidly now. Of course, uncertainties remain. Perhaps the outflow in SDSS J0905 + 57 is, after all, powered by a supermassive black hole shining as an AGN. AGNs are notoriously fickle. Because black holes are so small, the brightness of AGNs can fluctuate dramatically on short timescales. It is therefore hard to rule out the possibility that an AGN 'lit up' the galaxy a few million years ago and launched the observed winds, only to fade away before we could see it. At present, we have Occam's razor (the principle of simplicity) as our guide, reminding us that we need not invoke anything other than the observed starlight to explain the winds. However, future observations by the authors could test this hypothesis definitively by building a statistical sample of galaxies that have similar properties to those of SDSS J0905 + 57.

In any case, this object represents a remarkable new class of system, one with violent, massive outflows of dense molecular gas. Geach and colleagues caught a galaxy in the act of 'fighting back' against gravity, and forever changing its own future. ■

Philip F. Hopkins is at the California Institute of Technology, Pasadena, California 91125, USA.
e-mail: phopkins@caltech.edu

1. Geach, J. E. *et al.* *Nature* **516**, 68–70 (2014).
2. Dekel, A. & Silk, J. *Astrophys. J.* **303**, 39–55 (1986).
3. Croton, J. *et al.* *Mon. Not. R. Astron. Soc.* **365**, 11–28 (2006).
4. Hopkins, P. F. *et al.* *Astrophys. J.* **625**, L71–L74 (2005).
5. Heckman, T. M., Armus, L. & Miley, G. K. *Astrophys. J. Suppl.* **74**, 833–868 (1990).
6. Murray, N., Quataert, E. & Thompson, T. A. *Astrophys. J.* **618**, 569–585 (2005).

A molecular knife to dice depression

Chronic stress can cause depression in some individuals, but leaves others untouched. Engagement of a molecular pathway controlling the production of tiny RNA snippets might help to explain the difference. [SEE ARTICLE P.51](#)

GERHARD SCHRATT

Stress and adversity can put the balance of our lives at risk — the death of a loved one, loss of a job or other financial or personal problems can have detrimental effects on the human psyche. Although some people become depressed in such circumstances, others are more resilient and can successfully cope with difficult experiences. Resilience was first documented in the 1970s, but the molecular underpinnings of this phenomenon are still mostly elusive¹. In this issue, Dias *et al.*² (page 51) report that behavioural resilience in mice is promoted by the β -catenin protein, through control of the enzyme Dicer1.

Research over the past decade shows that a complex interplay between genetic and environmental factors underlies mood disorders, and points to some key molecular pathways involved in such conditions³. This work has also led to the development of antidepressants, but such drugs are effective in only about

two-thirds of people with depression⁴. A better understanding of the molecular pathways active in resilient individuals could inform us of alternative treatment strategies⁵.

β -Catenin controls brain function⁶, and molecules regulating β -catenin activity have been implicated in depression⁷. Dias and colleagues analysed the role of β -catenin in resilience, using a well-established mouse model of depression — chronic social defeat stress (CSDS)⁸. Briefly, male mice are repeatedly exposed to males from a physically superior strain. The encounter regularly ends with the defeat of the test mice. Although the fights do not usually result in severe physical injury, the repeated defeat leaves psychological scars that ultimately manifest in depressive behaviour. For example, mice that have experienced CSDS will largely avoid social contact, a symptom also observed in people with depression⁹.

The authors focused on a specific brain region, the nucleus accumbens (NAc), which is part of the forebrain. Although the NAc is

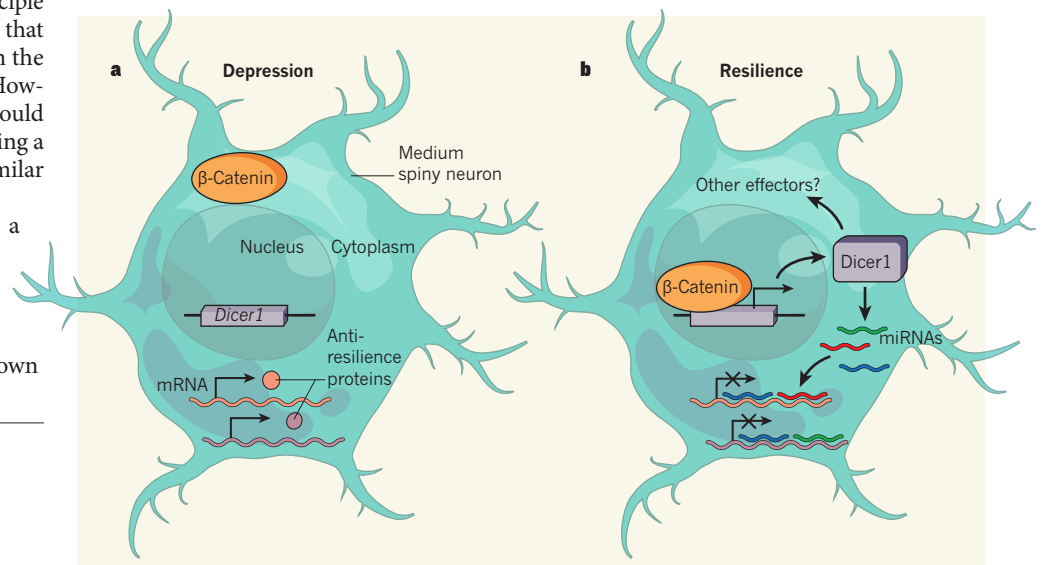


Figure 1 | A resilience switch. **a**, Dias *et al.*² report that D2-type medium spiny neurons are not activated in 'depressed' mice. As a consequence, β -catenin protein remains in the cytoplasm in these cells, unable to enter the nucleus, and the Dicer1 gene is thus inactive. 'Anti-resilience' proteins may therefore be produced from messenger RNA that would otherwise have been inhibited by microRNAs (miRNAs) generated by the Dicer1 protein. **b**, In resilient mice, β -catenin enters the nucleus of activated neurons, thereby turning on Dicer1 transcription. Elevated levels of Dicer1 protein increase production of miRNAs and possibly other effectors of resilience. This might, in turn, inhibit the production of anti-resilience proteins, because of binding and inhibition of mRNA by miRNAs.

primarily known for its role in addiction¹⁰, earlier work from the same group suggests that it is also involved in depression⁷. When Dias and co-workers artificially increased the amount of β -catenin in the NAc, the social behaviour of mice that had experienced CSDS became indistinguishable from that of the controls, suggesting that β -catenin promotes resilience. Conversely, blocking β -catenin resulted in depressive behaviour in mice that had experienced a stress dose that normally does not cause depression. Together, these experiments point to a pivotal role for β -catenin in stress resilience. Further investigation revealed that this action of β -catenin took place in only one type of neuron in the NAc, the D2-type medium spiny neurons, which are inactive in 'depressed' mice, but activated in those that show resilience.

These straightforward findings became more complicated when the authors tried to address exactly how β -catenin mediates pro-resilient behaviour. β -Catenin provides a classic example of 'protein moonlighting', a phenomenon in which a protein performs more than one function. It was originally identified as a component of a cell-adhesion complex¹¹ that, among other roles, regulates plasticity in mammalian neurons. But β -catenin also has a nuclear function as an intracellular signal transducer, relaying information from a class of extracellular signalling molecules, Wnts, to regulate gene expression in the nucleus¹². Dias *et al.* provide strong evidence that it is the nuclear function of β -catenin that is involved in resilience.

To gain insight into the genes regulated by β -catenin, the researchers created a genome-wide map of genetic regions bound by the protein. One of the most highly ranking genes to surface from this analysis was *Dicer1*, which encodes an enzyme that is crucial for the generation of various small regulatory RNAs¹³ — molecules that do not code for proteins themselves, but rather control protein production. The best-studied family of such RNAs, microRNAs (miRNAs), fulfil various functions in neurons¹⁴. Through a series of genetic experiments, the authors convincingly show that the pro-resilient effects of β -catenin are absent if *Dicer1* is genetically removed.

Dias *et al.* propose a model whereby β -catenin-dependent regulation of small RNAs — in particular miRNAs — underlies behavioural resilience to stress (Fig. 1). The potential link between β -catenin and miRNAs is perhaps the most exciting aspect of this study. Because each miRNA can regulate the production of hundreds of proteins, β -catenin could in this way control the expression of a plethora of 'anti-resilience' proteins that counteract resilience. However, several caveats remain to be addressed.

Dicer1 protein is necessary for the expression of most miRNAs, but only a few miRNAs were robustly affected by β -catenin

manipulation. Could this mean that other targets of *Dicer1* are involved in resilience? A growing body of work¹³ indicates that *Dicer1* is responsible for the production of other classes of small RNA (such as small interfering RNAs) and that it is also involved in small-RNA-independent processes. Comparing Dias and colleagues' results with those from other mice with reduced miRNA processing — for example, mice with mutations in the *Dgcr8* gene¹⁵ — should help to clarify the role of miRNAs in resilience. Even if miRNAs are the crucial effectors of resilience, identifying the physiologically relevant ones among the nearly 1,000 miRNA species seems to be a long shot. Nevertheless, Dias *et al.* provide some promising leads for follow-up experiments.

What is the relevance of these findings for humans? The β -catenin pathway is not unknown territory for therapeutic depression research. Several proteins that control β -catenin activity are altered in people with depression, including the kinase enzyme GSK3, which is a target of the antidepressant drug lithium⁴. It will now be necessary to determine if the β -catenin-regulated miRNAs presented in the current study are also affected in the brains of people with depression.

Dias and colleagues' research underscores the importance of small-RNA regulation in neuropsychiatric conditions. Assuming that effective and safe delivery of RNA into the human brain can be accomplished in the future, boosting the expression of selected

small RNAs could become a viable strategy for building a protective shield that prevents stress from throwing our minds off balance. ■

Gerhard Schratt is at the Biochemisch-Pharmakologisches Centrum, Institut für Physiologische Chemie, Philipps-Universität Marburg, 35032 Marburg, Germany. e-mail: schratt@staff.uni-marburg.de

1. Garmezy, N. in *Schizophrenia: The First Ten Dean Award Lectures* (ed. Dean, S. R.) 163–204 (MSS Info. Corp., 1973).
2. Dias, C. *et al.* *Nature* **516**, 51–55 (2014).
3. Kupfer, D. J., Frank, E. & Phillips, M. L. *Lancet* **379**, 1045–1055 (2012).
4. Duman, R. S. & Voleti, B. *Trends Neurosci.* **35**, 47–56 (2012).
5. Southwick, S. M. & Charney, D. S. *Science* **338**, 79–82 (2012).
6. Maguschak, K. A. & Ressler, K. J. *Neuropharmacology* **62**, 78–88 (2012).
7. Wilkinson, M. B. *et al.* *J. Neurosci.* **31**, 9084–9092 (2011).
8. Golden, S. A., Covington, H. E. III, Berton, O. & Russo, S. J. *Nature Protocols* **6**, 1183–1191 (2011).
9. Trew, J. L. *Clin. Psychol. Rev.* **31**, 1156–1168 (2011).
10. Russo, S. J. *et al.* *Trends Neurosci.* **33**, 267–276 (2010).
11. Ozawa, M., Baribault, H. & Kemler, R. *EMBO J.* **8**, 1711–1717 (1989).
12. Behrens, J. *et al.* *Nature* **382**, 638–642 (1996).
13. Foulkes, W. D., Priest, J. R. & Duchaine, T. F. *Nature Rev. Cancer* **14**, 662–672 (2014).
14. Schratt, G. *Nature Rev. Neurosci.* **10**, 842–849 (2009).
15. Babiarz, J. E. *et al.* *RNA* **17**, 1489–1501 (2011).

This article was published online on 12 November 2014.

TECHNOLOGY

Ultrafast imaging takes on a new design

An imaging technique has been developed that can record non-repetitive ultrafast phenomena without strobe or flash illumination. The approach could find applications in biomedicine and security technologies. SEE LETTER P.74

BRIAN W. POGUE

In this issue, Gao and colleagues¹ (page 74) describe a new imaging technology that can capture time-evolving events at a speed of up to 100 billion frames per second without the need for specialized illumination methods. The technique involves a unique hardware design combined with spatial encoding of the image, to achieve something not possible before. The approach allows imaging of ultrafast phenomena that need happen only once, and does so using single camera snapshots. By contrast, ultrafast imaging techniques used today rely on imaging the same subject over and over again, requiring it to be

a repeated event, or on using strobe or flash illumination.

Imaging fast phenomena has been a research field in itself, ever since the golden age of film photography in the nineteenth century, which used controllable mechanical light shutters that limit the film exposure time. High-speed photography using moving prisms or mirrors was originally developed in the twentieth century, and has been used to take spectacular photos of bullets going through apples and water balloons, to capture suspended water droplets, explosion clouds and sonic booms, or simply to record sporting events. The development of digital electronic sensors in the semiconductor era of the

1970s to '80s completely changed high-speed photography, allowing fast events to be imaged electronically (Fig. 1). Since then, there have been fundamental design changes to the methods used for high-speed exposures, allowing ever faster imaging. This has advanced to the point that ultrafast imaging today can observe the movement of light on the millimetre distance scale.

Mid-twentieth-century approaches typically captured phenomena on the millisecond to microsecond timescale. They required highly efficient sensors (such as charge-coupled devices and complementary metal-oxide-semiconductor sensors), sufficient lighting, or strobe flash photography. Investigation into the scientific origins of fast phenomena has benefited enormously from the development of such high-speed cameras, and other applications of this technology exist in all areas of human interaction, commerce, health care and defence. The key technological leap to imaging at timescales smaller than one microsecond came with the invention of advanced rotating-prism cameras and rotating-mirror cameras.

However, imaging of ultrafast phenomena requires something extra special, and the applications of ultrafast imaging can be even more compelling than those of fast imaging. The main breakthrough in this technology capitalized on the proven idea that, by moving the sensor lateral to the direction of imaging, a temporal signal could be captured spatially on the sensor. In this way, time variations could be translated into spatial variations on the sensor. This conversion from time to space could be done with a moving mirror, a film sequence or a moving sensor. However, the major advance came through converting the temporal signal into electrons at a photocathode inside a vacuum tube, amplifying the converted signal and spreading it laterally through electronics. This architecture became known as a streak camera, for its ability to turn a time-varying signal into a horizontal 'streak of light' on an electronic image sensor. For decades, this technique has been used as a commercially available ultra-high-speed imaging approach, in which images are read out by a standard electronic camera.

Since the 1980s, electronic-amplification vacuum tubes known as gated microchannel plate photomultipliers have been the core technology in streak cameras, and have allowed imaging with picosecond (10^{-12} s) temporal resolution and less than nanosecond-level shutter times. However, their design typically limited their use to a single, one-dimensional data set, with the second dimension being used to spread the data temporally before amplification in the photomultipliers. With this design, the output image displays horizontal position versus time, but image-acquisition rates are less than 1 billion frames per second.

Techniques to reach imaging in the billions

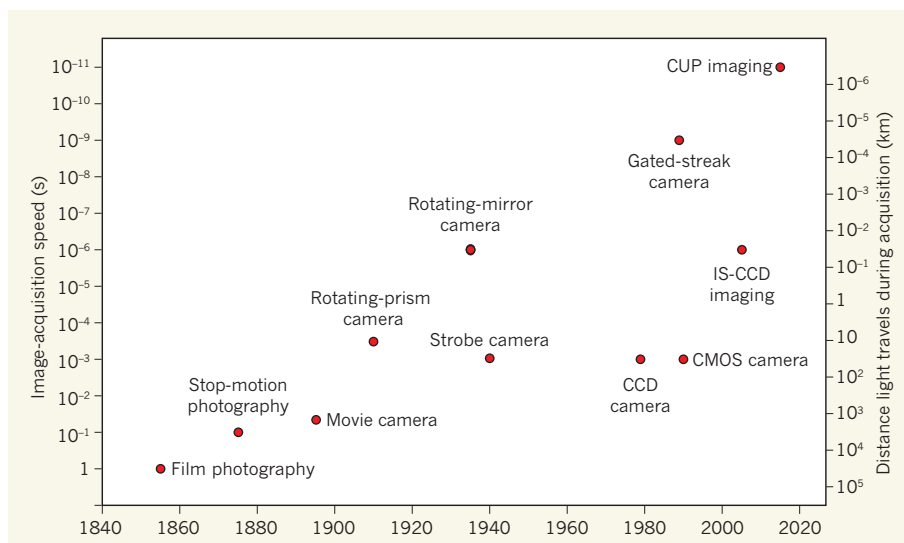


Figure 1 | Fast cameras over time. Imaging at slow acquisition speeds has its origins in the nineteenth century, with fast imaging using rotating prisms, mirrors or strobe cameras arriving in the twentieth century. Modern digital cameras based on charge-coupled devices (CCDs) and complementary metal-oxide-semiconductor (CMOS) sensors were developed in the 1980s. These sensors then enabled electronic technologies to be developed for ultrafast imaging, such as gated-streak cameras and *in situ* storage CCD (IS-CCD) devices. Gao *et al.*¹ have developed a technique called compressed ultrafast photography (CUP), which outperforms previous ultrafast imaging technologies without the need for active illumination.

of frames per second typically need a radical new design. One such advance that emerged² in 2009 involved encoding each of the pixels in an image in spectral wavelengths of light, and then transforming this spectrum into a time sequence of data. This allowed the image pixels to be serially amplified by a fibre laser and read out, pixel by pixel, by a single detector. This imaging infrastructure does not look like a camera at all, but the image could be created pixel by pixel, with a frame rate of around 6 million frames per second, and an effective light-exposure time of less than 0.5 ns. The major benefit of this approach was that the amplification of the signal above a factor of 300 provided a superior resolution, and so the technique could be applied to low-intensity phenomena. However, the infrastructure required to create the images was high.

In comparison, Gao and colleagues' work shows that a more conventional set-up of a camera allows imaging at the extremely high frequency of 100 billion frames per second. The authors' method, called compressed ultrafast photography, uses the imaging optics and image geometry of streak cameras, but takes advantage of the ability of compressed sensing tools to recover images from sparse spatial data. The approach enables the information in the image to be laterally encoded and also somewhat randomly encoded in the field of view of the streak camera. This encoded-image data set allows the system to read out images with full-frame capability at extremely high speed.

So, what could be done with an ultrafast camera at 10^{11} frames per second? Applications could include visualizing optical

communications, optically active light-matter interactions and quantum-mechanical phenomena. For example, it might be possible to improve the investigation into approaches to optical cloaking³, in which light bends or is deformed around an object, instead of going through it. This field of study, popularized in *Star Trek*, is real, and although many advances are being made in fundamental approaches to cloaking designs, the inability to see the interactions between light and the object being cloaked hampers development. Similarly, phenomena in which light can focus or defocus through a material, and effects in which light oscillates between thin layers, could all be imaged for the first time. It is also likely that new ways to image high-speed signals could result in innovations in industrial processing for materials research, biomedicine and security technologies.

The progression of fast-imaging technologies has been steady ever since the invention of film, but key transformative design innovations such as the one reported by Gao and colleagues are still few and far between. These advances will be essential for seeing the physical behaviour of light and exploiting it. ■

Brian W. Pogue is in the Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire 03755, USA.
e-mail: brian.w.pogue@dartmouth.edu

1. Gao, L., Liang, J., Li, C. & Wang, L. V. *Nature* **516**, 74–77 (2014).
2. Goda, K., Tsia, K. K. & Jalali, B. *Nature* **458**, 1145–1149 (2009).
3. Cai, W., Chettiar, U. K., Kildishev, A. V. & Shalae, V. M. *Nature Photon.* **1**, 224–227 (2007).

Professional identity can increase dishonesty

An experiment shows that although bank employees behave honestly on average, their dishonesty increases when they make decisions after having been primed to think about their professional identity. [SEE LETTER P.86](#)

MARIE CLAIRE VILLEVAL

Honesty involves individual preferences, but also legal rules and moral and social norms that prescribe what we ought to do. We would like to believe that such norms are able to help individuals resist the temptation of earning more money by behaving dishonestly. But unfortunately, norms that are specific to a particular location or group can sometimes have the opposite effect. On page 86 of this issue, Cohn *et al.*¹ provide an example of this in their report of an experiment conducted with the employees of a large international bank. The authors show that the bank employees behave honestly on average in a control situation, but become less honest after having been 'primed' to think about their professional identity.

Several scandals in the financial sector have tarnished its reputation, but it has been unclear whether these scandals result mainly from the effects of selection or of incentives (Fig. 1). The first possibility centres

on the idea that the considerable amounts of wealth and assets traded in the financial sector attract some people who have weaker honesty norms. According to the second explanation, the incentives and the business culture developed in the financial sector may undermine the honesty norms of ordinary employees. Cohn *et al.* provide evidence for the latter explanation.

A direct comparison of the behaviour of bankers and their respective degrees of honesty with that of employees from other sectors would not be conclusive because one can never guarantee that the two groups would be perfectly comparable, owing to unobservable characteristics. Instead, Cohn *et al.* separated their banker participants into two experimental groups. In the control group, the employees completed a survey about life satisfaction, well-being and everyday life that did not include questions related to their professional role. They were then asked to toss a coin ten times in privacy and report the outcomes online. For each toss, the participants

knew in advance whether tails or heads would yield a monetary pay-off. In the treatment group, seven of the questions in the preliminary questionnaire were replaced with those related to the bankers' profession. For example, questions such as "how many hours per week do you watch television on average?" were replaced with ones such as "what is your function at this bank?". These questions were designed to prime the participants to their professional identity before they completed the coin-toss and reporting task.

The authors found that, in the control group, the distribution of winning tosses reported did not differ from the binomial distribution expected from honest reporting. This indicates that, on average, the employees behaved honestly. However, reporting by the treatment group was significantly different: the percentage of winning tosses reported increased from 51.6% in the control group to 58.2% in the treatment group, and the fraction of misreported coin flips increased from 3% to 16%.

The effect induced by the treatment could be attributable to several causes, including the competitiveness expected from bank employees, the exposure to competitive bonus schemes, the beliefs about what other employees would do in the same situation or the salience of money in the questionnaire. Cohn and colleagues conducted further tests and statistical analyses to assess these potential channels, but provide data that discount each of them. They also find that the professional-identity priming effect was not replicated in identical sessions involving participants from other professions. This strongly suggests that the increased dishonesty of the bank employees in the treatment group results from the business culture specific to the banking industry. Business culture is a complex entity that encompasses an organization's norms, values, visions, expectations and habits; it largely exceeds the incentives schemes promoted in the company.

These findings confirm some popular opinions about practices in the financial sector and they have direct implications: it is crucial to ensure a business culture of honesty in this industry to restore trust in it. Testing whether the professional-identity effect disappears in companies that have introduced training programmes in ethics would be extremely relevant. The results also open the door to other possible tests of the professional-identity effect on preferences. For example, risk-taking decisions may also be strongly affected by professional-identity priming. Moreover, there is no reason to believe that this effect is limited to the financial sector. For example, the same method could be used to test whether the honesty of politicians is negatively affected by political-background priming when the participants are faced with opportunities for political gain, rather than the financial incentive used for the bankers. At the opposite end of the moral-norms spectrum, the degree of



Figure 1 | Primed by the profession. The financial sector has been tarnished by several scandals, such as that involving American stockbroker and financier Bernard Madoff, who was convicted in 2009 in relation to billions of dollars of fraudulent investments. Cohn and colleagues' experimental findings¹ suggest that the business culture of the financial industry can undermine the honesty norm of its employees.

altruism of doctors and nurses or the willingness of police officers to enforce punishment might increase when primed with their professional identity.

From a scientific perspective, this study not only supports the economic theory of social identity^{2–5} — which considers that gender or occupation, for example, create multiple social identities in individuals — but also links this theory with the economic analysis of lying behaviour^{6,7}. Beyond economics, it shows how behavioural economists can contribute to a broader reflection in science about how people manage their ‘multiple selves’ to maintain their self-concept of honesty⁸. In particular, we should explore further whether professional identity is used strategically by people to reduce their own responsibility for their behaviour and to shift the blame to their companies. Neuroscientific methods could be solicited to understand how and when the human brain lets professional (impersonal) identity become the dominant

influence on an individual’s moral norms.

It was not Cohn and colleagues’ aim to explain how business culture may encourage misbehaviour. However, understanding how the culture of dishonesty evolves is an important issue and it is unlikely that the process is one way. On the one hand, the development of business culture may lead to a ‘deindividuation’ process that facilitates the alignment of the individual with the group’s preferences⁹. If the business culture goes wrong, then individuals may also develop unethical behaviour. On the other hand, individual misconduct may also influence the evolution of business cultures towards more deviant collective norms, through social influence and belief formation. A better understanding of this dual process will require a theoretical framework that takes into account not only individual motivations but also social interactions and peer effects¹⁰. ■

Marie Claire Villeval is at GATE, Centre

National de la Recherche Scientifique and University of Lyon, 69007 Lyon, France. She is also a member of the CORTEX Laboratory of Excellence.

e-mail: villeval@gate.cnrs.fr

1. Cohn, A., Fehr, E. & Maréchal, M. A. *Nature* **516**, 86–89 (2014).
2. Akerlof, G. A. & Kranton, R. E. *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being* (Princeton Univ. Press, 2010).
3. Chen, Y. & Li, S. X. *Am. Econ. Rev.* **99**, 431–457 (2006).
4. Goette, L., Huffman, D. & Meier, S. *Am. Econ. Rev.* **96**, 212–216 (2006).
5. Benjamin, D. J., Choi, J. J. & Strickland, A. J. *Am. Econ. Rev.* **100**, 1913–1928 (2010).
6. Fischbacher, U. & Föllmi-Heusi, F. J. *Eur. Econ. Assoc.* **11**, 525–547 (2013).
7. Gneezy, U. *Am. Econ. Rev.* **95**, 384–394 (2005).
8. Ariely, D. *The Honest Truth About Dishonesty* (Harper Collins, 2012).
9. Reicher, S. D., Spears, R. & Postmes, T. *Eur. Rev. Soc. Psychol.* **6**, 161–198 (1995).
10. Lefebvre, M., Pestieau, P., Riedl, A. & Villeval, M. C. *Int. Tax Public Finance* <http://dx.doi.org/10.1007/s10797-014-9318-z> (2014).

This article was published online on 19 November 2014.

DIABETES

The good in fat

A new class of fatty acid — found in food and synthesized by mammalian tissues — enhances glucose uptake from the blood and reduces inflammation, suggesting that these fats might be used to treat diabetes.

DEBORAH M. MUOIO &
CHRISTOPHER B. NEWGARD

The word fat has an almost universally negative connotation in modern society, whether as a pejorative term for describing someone who is overweight or as a foodstuff reviled for its deleterious impact on health. Obesity is associated with elevated levels of multiple kinds of fat, and fat accumulation in tissues is thought to contribute to the development of the unholy trinity of disorders that characterize type 2 diabetes: failure to respond to the glucose-regulating hormone insulin; failure to control glucose production by the liver; and failure of pancreatic β -cells, which store and release insulin. Now, however, Yore *et al.*¹ report in *Cell* that certain modified fatty acids are found at low, rather than elevated, levels in obese, insulin-resistant humans and mice. Furthermore, they show that administering these fatty acids to such mice improves glucose uptake from the blood, enhances insulin secretion and relieves obesity-associated inflammation, suggesting that these naturally occurring fats could be used for diabetes therapy.

Fatty acids and closely related glycerolipid molecules wear multiple biological hats, serving as a major energy reservoir, as constituents

of cell membranes and as mediators of intracellular signalling. Yore and colleagues’ report adds to emerging evidence implicating specific, low-abundance fatty acids as hormone-like ‘lipokine’ molecules involved in metabolic regulation.

The same research group had previously shown² that elimination of the glucose-transport protein GLUT-4 from adipose (fat-storing) tissue resulted in insulin resistance in liver and muscle — a condition in which cells of these organs have an impaired capacity to increase glucose uptake in response to insulin, contributing to the abnormally high blood-glucose levels that define diabetes. By contrast, adipose-specific overexpression of GLUT-4 caused higher glucose uptake than in control mice — making the mice more sensitive to insulin despite also increasing fat synthesis and levels of circulating fatty acids³.

The lipid-generating impact of GLUT-4 overexpression was predictable, given that fat cells (adipocytes) use glucose for fat synthesis. But surprisingly, the enhanced glucose uptake was eliminated by adipose-specific deletion of the transcription factor ChREBP, which regulates fat synthesis, suggesting that the glucose-uptake effect depended on lipogenesis⁴. This finding led Yore and colleagues to the hypothesis that certain fatty acids have

positive effects on glucose regulation.

The authors used mass spectrometry to compare the lipid content of the blood serum and adipose tissue taken from these ‘diabetes-resistant’ mice to that of normal mice. They identified five lipids that were significantly elevated in the modified mice, four of which were made up of a typical long-chain fatty acid (palmitate, oleate, stearate or palmitoleate) joined by an ester bond to a hydroxylated version of one of the same set of fatty acids. The authors refer to these compounds, which had not been described previously, as fatty-acid esters of hydroxyl fatty acids (FAHFAs).

Interestingly, only 6 of the possible 16 FAHFA species were upregulated by GLUT-4 overexpression. The most dramatically upregulated type of FAHFA in the GLUT-4-overexpressing mice was palmitic acid hydroxystearic acid (PAHSA), which was then studied further. Yore *et al.* report that PAHSA (which exists in several isomeric forms depending on the esterification position) can be synthesized by mice provided with the hydroxylated fatty-acid precursors, and that it and other FAHFAs are present in mouse serum at low nanomolar concentrations and in laboratory-mouse and human food (Fig. 1). Surprisingly, the levels of FAHFAs are much higher in low-fat, ‘standard’ mouse diets than in high-fat mouse diets used to induce obesity and insulin resistance. These findings suggest that both dietary intake and endogenous synthesis can influence tissue and blood concentrations of FAHFAs.

The authors then studied these fatty acids in two models of insulin resistance in mice: fasting and high-fat feeding. Although they found that PAHSA levels fluctuate in response to nutritional state, the direction and magnitude of the changes depended on the PAHSA isoform analysed and the tissue in which

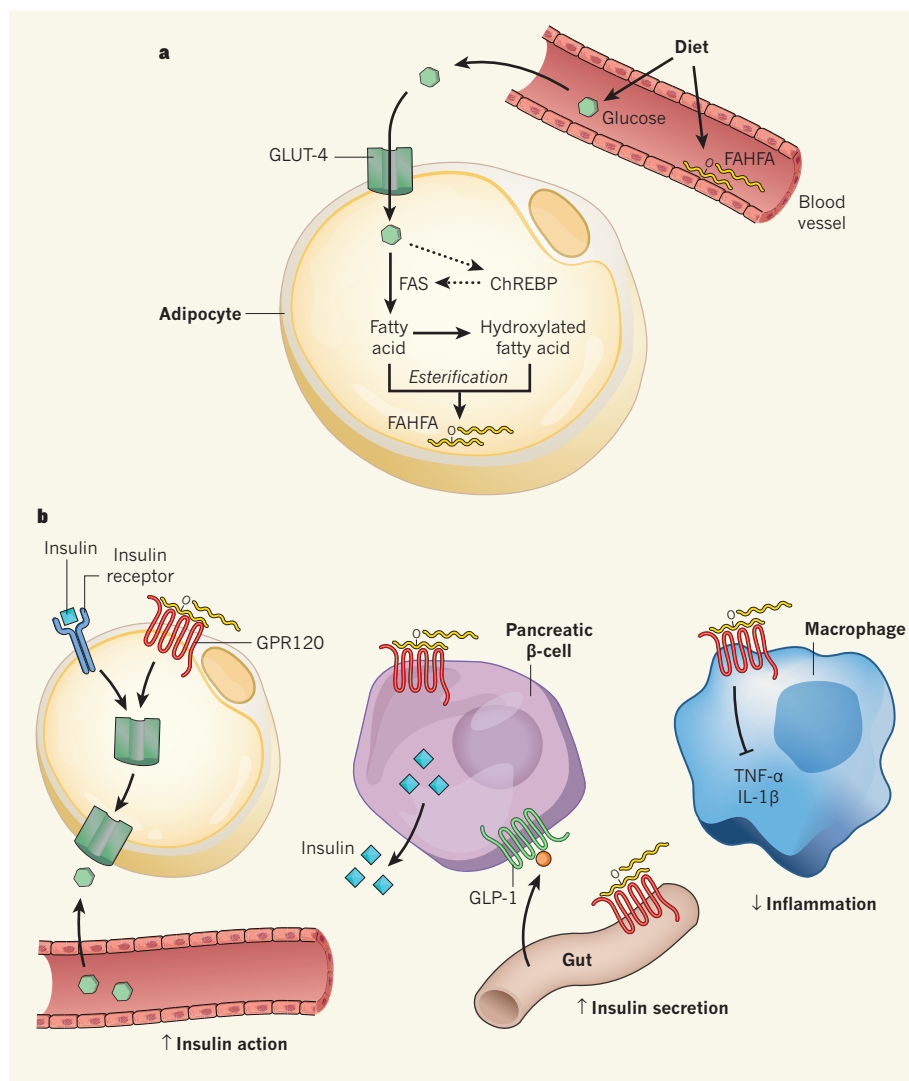


Figure 1 | Anti-diabetic and anti-inflammatory effects of fatty acids. **a**, Yore *et al.*¹ describe a new class of fatty acid, called fatty-acid esters of hydroxyl fatty acids (FAHFAs), that are present in foods and that can be produced by adipocytes from newly synthesized long-chain fatty acids that are joined by esterification to hydroxylated fatty acids. Lipogenesis, the process of converting glucose to new fats, is catalysed by the enzyme FAS and activated by ChREBP, a glucose-responsive transcription factor. **b**, The authors show that administering FAHFAs to mice results in more-efficient clearance of glucose from the blood and suggest that this occurs through increasing secretion of, and sensitivity to, insulin, a hormone secreted by pancreatic β -cells. In adipocytes *in vitro*, FAHFAs activate the receptor GPR120, which increases translocation of the glucose receptor GLUT-4, and thereby enhances insulin-stimulated glucose uptake. Administration of FAHFAs *in vivo* leads to increased serum levels of insulin and glucagon-like peptide-1 (GLP-1), a gut hormone that enhances insulin secretion. In addition to these anti-diabetic effects, FAHFA activation of GPR120 reduces the production of the inflammatory molecules IL-1 β and TNF- α by macrophage cells of the immune system.

they were measured, suggesting that PAHSA abundance in various tissues is not clearly correlated with insulin sensitivity in these rodent models. However, in humans, they observed a strong association between PAHSA and insulin sensitivity, and PAHSA levels were significantly lower in insulin-resistant than in insulin-sensitive individuals.

To test the functional significance of PAHSA in glucose regulation, the authors administered the fatty acid orally to mice before performing a glucose-tolerance test, which measures how quickly a dose of glucose is cleared from the

blood. They observed reduced basal glucose levels and more-efficient glucose clearance than in mice that had not received PAHSA, and suggested that this effect might be due to improved insulin sensitivity. Although they did not test this idea in mice, they observed a modest potentiation of insulin-stimulated glucose uptake in adipocytes *in vitro* following the addition of PAHSA. However, PAHSA administration to mice also caused increased levels of insulin and glucagon-like peptide-1 (a gut hormone that enhances insulin secretion) in the serum, which might also explain

the improved glucose uptake (Fig. 1). Thus, the mechanism by which PAHSA lowers blood glucose remains to be defined.

FAHFAs are not the only lipids for which therapeutic roles have been suggested in diabetes and related metabolic disorders. For example, administration of a monomeric fatty acid, palmitoleate, which is also produced in adipose tissue during periods of increased fat synthesis, decreases the liver fat content of mice and enhances insulin sensitivity in their skeletal muscles⁵. However, the strong associations between PAHSA and human insulin resistance demonstrated by Yore *et al.* have not been seen for palmitoleate. Omega-3 fatty acids have been suggested to protect against metabolic diseases by stimulation of the G-protein-coupled receptor GPR120, resulting in reduced inflammation, improved insulin sensitivity and translocation of GLUT-4 from the cell interior to the cell surface⁶. Yore *et al.* demonstrate that PAHSA also decreases the production of the pro-inflammatory molecules IL-1 β and TNF- α by macrophage cells, and that both this effect and PAHSA-induced enhancement of glucose uptake and GLUT-4 translocation are prevented by inhibiting the activity of GPR120 (Fig. 1). Thus, it seems that FAHFAs and omega-3 fatty acids may converge, at least in part, on the same receptor system to regulate glucose uptake and inflammation. It will be interesting for future studies to compare which is the more potent and efficacious lipid in this regard.

It remains to be determined how the different actions of FAHFAs — sensitizing cells to insulin, stimulating insulin production and reducing inflammation — contribute to glucose homeostasis. Other questions worthy of future investigation include which enzymes are involved in synthesizing and degrading FAHFAs and how they are regulated by changes in diet or, perhaps, by the composition of microorganisms in the gut. Finally, and importantly, the effects of FAHFA administration on glucose control in humans with type 2 diabetes remains to be explored. Results of these studies will determine whether FAHFAs or their derivatives may indeed become therapeutic agents for metabolic diseases. ■

Deborah M. Muoio and Christopher B. Newgard are in the Sarah W. Stedman Nutrition and Metabolism Center and Duke Molecular Physiology Institute, and the Departments of Pharmacology and Cancer Biology and Medicine, Duke University Medical Center, Durham, North Carolina 27701, USA. e-mail: chris.newgard@duke.edu

1. Yore, M. M. *et al.* *Cell* **159**, 318–332 (2014).
2. Abel, E. D. *et al.* *Nature* **409**, 729–733 (2001).
3. Shepherd, P. R. *et al.* *J. Biol. Chem.* **268**, 22243–22246 (1993).
4. Herman, M. A. *et al.* *Nature* **484**, 333–338 (2012).
5. Cao, H. *et al.* *Cell* **134**, 933–944 (2008).
6. Oh, D. Y. *et al.* *Cell* **142**, 687–698 (2010).

This article was published online on 19 November 2014.

β -catenin mediates stress resilience through *Dicer1*/microRNA regulation

Caroline Dias^{1*}, Jian Feng^{1*}, Haosheng Sun¹, Ning yi Shao¹, Michelle S. Mazei-Robison^{1†}, Diane Damez-Werno¹, Kimberly Scobie¹, Rosemary Bagot¹, Benoit LaBonté¹, Efrain Ribeiro¹, XiaoChuan Liu¹, Pamela Kennedy^{1†}, Vincent Vialou^{1†}, Deveroux Ferguson^{1†}, Catherine Peña¹, Erin S. Calipari¹, Ja Wook Koo¹, Ezekiel Mouzon¹, Subroto Ghose², Carol Tamminga², Rachael Neve³, Li Shen¹ & Eric J. Nestler¹

β -catenin is a multi-functional protein that has an important role in the mature central nervous system; its dysfunction has been implicated in several neuropsychiatric disorders, including depression. Here we show that in mice β -catenin mediates pro-resilient and anxiolytic effects in the nucleus accumbens, a key brain reward region, an effect mediated by D2-type medium spiny neurons. Using genome-wide β -catenin enrichment mapping, we identify *Dicer1*—important in small RNA (for example, microRNA) biogenesis—as a β -catenin target gene that mediates resilience. Small RNA profiling after excising β -catenin from nucleus accumbens in the context of chronic stress reveals β -catenin-dependent microRNA regulation associated with resilience. Together, these findings establish β -catenin as a critical regulator in the development of behavioural resilience, activating a network that includes *Dicer1* and downstream microRNAs. We thus present a foundation for the development of novel therapeutic targets to promote stress resilience.

Despite decades of research, the molecular pathophysiology of depression remains elusive. One molecular player implicated in neuropsychiatric illnesses, including depression, is β -catenin^{1–5}. In addition to having a structural role at synapses, β -catenin mediates the transcriptional output of canonical Wnt signalling^{6–8}. This multi-functionality has made it difficult to untangle the mechanism through which β -catenin might contribute to pathological states. We recently demonstrated the involvement of upstream Wnt signalling in the nucleus accumbens (NAc) in mouse depression models, with impaired signalling mediating susceptibility to social stress and increased signalling mediating resilience⁹. We thus began by studying the behavioural role of β -catenin in this brain region.

β -catenin mediates resilience and anxiolytic responses

We overexpressed β -catenin in a herpes simplex virus (HSV) vector in NAc (Fig. 1a; Extended Data Fig. 1a), which increases β -catenin solely in the nuclear compartment, as measured by subcellular fractionation and immunohistochemistry (IHC), whereas global N-cadherin/ β -catenin complexes were unaffected (Extended Data Fig. 1b, c). This suggests that HSV- β -catenin selectively activates the transcriptional function of the protein, without having direct effects on N-cadherin at synapses, consistent with earlier work in cultured cells¹⁰.

We next overexpressed β -catenin in NAc during accelerated social defeat stress (ASD)^{11,12}. We found that, while HSV-GFP injected control animals developed social avoidance, an indicator of depression-like behaviour, overexpression of β -catenin prevented this phenotype (Fig. 1b). Furthermore, in baseline behavioural assays, β -catenin mediated an antidepressant-like response in the forced swim test (FST) (Fig. 1c), and anxiolytic effects in the elevated plus maze (EPM) (Fig. 1d). We saw no changes in sucrose preference or cocaine conditioned place preference (data not shown), suggesting that β -catenin does not cause hedonic

changes. To confirm the pro-resilient effect of β -catenin, we used a stabilized β -catenin mutant (S33Y)¹³, and found identical results for wild-type β -catenin in the ASD and FST (Supplementary Notes), with no change in sucrose preference (data not shown). Finally, cell-type-specific overexpression of β -catenin in D2- but not D1-type medium spiny neurons (MSNs) in NAc (Fig. 1e, Extended Data Fig. 2a) induced a pro-resilient phenotype.

We also investigated the consequences of blocking β -catenin signalling in NAc with two approaches: excising β -catenin from NAc of conditional floxed mice (Extended Data Fig. 2b) and overexpressing a behaviourally validated dominant negative β -catenin mutant (Extended Data Fig. 2c)¹⁴. Both manipulations promoted susceptibility to stress in mice subjected to a sub-threshold defeat procedure (Fig. 1f, g). Excising β -catenin from NAc caused no change in social interaction or locomotion in control animals, demonstrating a specific association with stress (Extended Data Fig. 3a–c). These results establish a critical role for β -catenin signalling in NAc in behavioural resilience.

To explore the endogenous activity of β -catenin in depression, we examined its transcriptional activity in post-mortem NAc of depressed humans. *Axin2*, a universal readout of activated canonical β -catenin signalling, was robustly suppressed in NAc of depressed humans (Fig. 2a, Supplementary Table 1, Extended Data Fig. 4a). In contrast, total N-cadherin and β -catenin messenger RNA levels were unchanged, pointing specifically to β -catenin nuclear function alterations in depression. There was also suppression of *Tcf3* and *Tcf4* (T cell transcription factors 3 and 4) levels in depressed patients (Fig. 2a); these are two of several transcription factors through which β -catenin acts. Together, these data demonstrate downregulation of the transcriptional output of β -catenin in NAc in human depression.

We next investigated *Axin2* mRNA levels in mouse NAc 48 h after chronic social defeat stress (CSDS). We found no difference between

¹Fishberg Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ²Department of Psychiatry, University of Texas Southwestern, Dallas, Texas 75390, USA. ³Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [†]Present addresses: Department of Physiology, Michigan State University, East Lansing, Michigan 48824, USA (M.S.M.-R.); Department of Psychology, UCLA College of Life Sciences, Los Angeles, California 90095, USA (P.K.); Institut National de la Santé et de la Recherche Médicale (INSERM) U1130; CNRS UMR8246; UPMC UMR18, Neuroscience Paris Seine, 75005 Paris, France (V.V.); Department of Basic Medical Sciences, The University of Arizona College of Medicine-Phoenix, Arizona 85004, USA (D.F.).

*These authors contributed equally to this work.

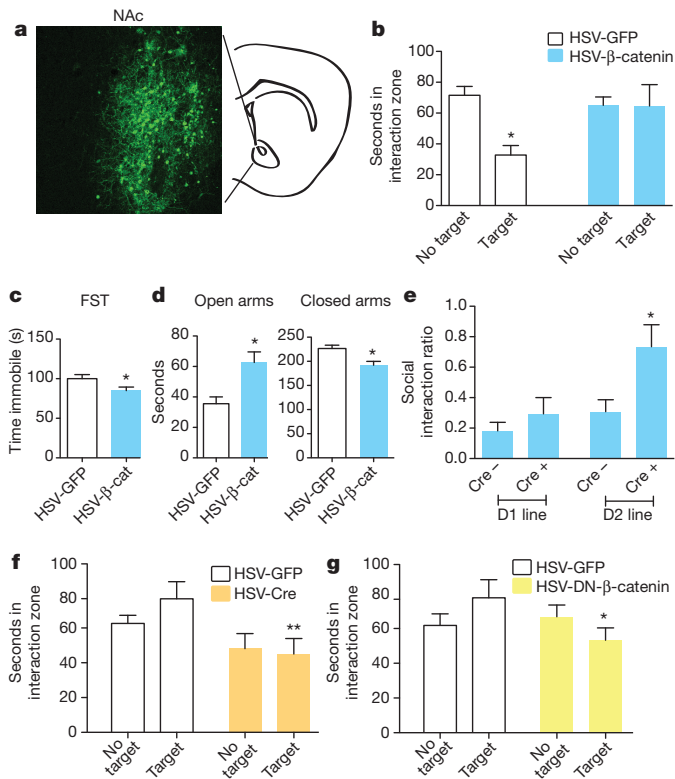


Figure 1 | β -catenin in NAc mediates pro-resilient, antidepressant, and anxiolytic responses. **a**, IHC illustrating viral transgene expression mediated by HSV- β -catenin with coronal cartoon of NAc highlighted. **b**, Pro-resilient effect of HSV- β -catenin on social interaction after ASD (* $P < 0.05$, two way ANOVA, $n = 8$ GFP, $n = 10$ β -catenin). **c**, Antidepressant-like effect of β -catenin in the forced swim test (* $P < 0.05$, two-tailed t -test, $n = 6$ GFP, $n = 7$ β -catenin). **d**, Anxiolytic-like effect of β -catenin in the elevated plus maze (closed arms: * $P < 0.01$, open arms: * $P < 0.01$, two-tailed t -test, $n = 6$ GFP, $n = 7$ β -catenin). **e**, Cell-type-specific overexpression of β -catenin in ASD (D2 Cre⁻ versus Cre⁺: * $P < 0.05$, two-tailed t -test, $n = 13$ D2 Cre⁻, $n = 8$ D2 Cre⁺). **f**, Effect of knocking down β -catenin in a sub-threshold defeat procedure (** $P < 0.01$, two-way ANOVA, effect of virus only when target present, $n = 6$ GFP, $n = 5$ Cre). **g**, Effect of dominant negative β -catenin in sub-threshold defeat (* $P < 0.05$, two-way ANOVA, interaction effect, $n = 5$ GFP, $n = 4$ dominant negative). Data presented as mean and s.e.m. and are representative of at least two experiments. See Methods and Supplementary Table 9 for detailed statistics.

susceptible and resilient animals (Fig. 2b). However, resilient animals displayed increased Tcf3 and Tcf4, indicating that resilience may be associated with upregulation of β -catenin signalling (Fig. 2b). To probe this, we examined the levels of phospho-Ser 675 β -catenin, a form with enhanced transcriptional activity, as well as total β -catenin at this time point. We found upregulation in resilient versus susceptible animals of phospho-Ser 675 β -catenin but not total β -catenin (Extended Data Fig. 4b). At 10 days after CSDS, we found elevated levels of Axin2 in resilient mice only ($P < 0.05$, Supplementary Notes).

Cell-type-specific action of β -catenin in resilience

Given the small magnitude of change observed above, we questioned whether the cell-type-specific behavioural effects in Fig. 1e corresponded to differential regulation of β -catenin signalling in D2 versus D1 MSNs. Using fluorescence-assisted cell sorting-isolated NAc neurons from D2-GFP mice (whereby the D2 neurons are labelled with green fluorescent protein, GFP), we found robust induction of Axin2 expression in D2⁺ neurons of resilient mice, and significantly reduced Axin2 levels in susceptible versus resilient mice, 48 h post CSDS, effects not seen in D2⁻ cells (Fig. 2c). Furthermore, Axin2 IHC with D1- or D2-GFP transgenic mice subjected to CSDS revealed downregulation of β -catenin

transcriptional activity in D2 versus D1 MSNs in susceptible mice (Fig. 2d). In sum, upregulation of β -catenin signalling occurs in D2 MSNs in resilient mice, with downregulation seen in susceptible animals.

Because glutamatergic neurotransmission regulates β -catenin transcriptional activity and stress susceptibility^{15,16}, we tested whether medial prefrontal cortex (PFC) or hippocampus, two important glutamatergic inputs to NAc, control β -catenin signalling in NAc. Using previously validated constructs and stimulation protocols^{17,18}, we found that optogenetic stimulation of glutamatergic PFC terminals robustly suppressed β -catenin activity in NAc as indicated by decreased Axin2, Tcf3, and Tcf4, whereas stimulation of hippocampus terminals had no effect (Fig. 2e, f). Repeated burst firing of dopamine afferents from the ventral tegmental area (VTA) also had no effect (Extended Data Fig. 5). Thus, PFC to NAc stimulation specifically elicited a molecular 'signature' of susceptibility, indicating that activation of this circuit could mediate the maladaptive suppression of β -catenin activity in NAc.

Genome-wide mapping of β -catenin after social defeat

We next conducted β -catenin chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) on NAc of control, susceptible, and resilient mice after CSDS. We first validated our β -catenin ChIP protocol by examining an LEF/TCF consensus sequence in the promoter of a known β -catenin target gene, *CaMKIV* (also known as *Camk4*). We found enrichment of β -catenin at the LEF/TCF site, but not a distant site, in NAc of resilient mice only (Fig. 3a). Through ChIP-seq^{19,20} we then examined global β -catenin enrichment after CSDS, and found major differences in peak numbers (Fig. 3b, Supplementary Data 1). Control and resilient conditions were associated with 10–15-fold higher absolute peak numbers compared to susceptible conditions, suggesting profound global alterations in β -catenin activity, consistent with our biochemical data (Fig. 2). Enrichment of β -catenin in resilient animals (Fig. 3b) only occurred at transcriptionally active sites, as indicated by high basal binding of two transcriptional activation marks H3K4me3 and H4K16ac (Fig. 3c, Extended Data Fig. 6). However, we did not observe global changes in these two histone marks after CSDS (Extended Data Figs 7, 8), suggesting that β -catenin may be recruited to active, open regions of chromatin through the presence of other, direct DNA-binding transcription factors.

Using Ingenuity pathway analysis, we demonstrated a predicted β -catenin network to be upregulated in NAc of resilient versus susceptible mice (Extended Data Fig. 9), a prediction specific to β -catenin. Concomitantly, there were nearly twice as many increases as decreases in β -catenin binding in resilient versus control mice at promoter regions. In contrast, susceptible versus control animals displayed equivalent numbers of upregulated and downregulated β -catenin binding events (Fig. 3d). These results support our hypothesis that resilience is associated with genome-wide enrichment of β -catenin. Examining the distribution of β -catenin peaks across the genome (Fig. 3e) revealed similar results: redistribution of β -catenin binding towards promoters and gene bodies in resilience, and redistribution away from promoters/gene bodies and towards gene deserts in susceptibility.

To validate the β -catenin ChIP-seq data, we conducted quantitative ChIP (qChIP) on independent biological samples at genes that showed significant peaks in resilience or upregulation in resilient versus susceptible animals, thus confirming significant β -catenin enrichment at several promoters (Fig. 3f). As further validation, we examined mRNA levels of genes found in our ChIP-seq list that coincided either with *in silico* lists of predicted or known β -catenin targets^{21,22} (Supplementary Table 2) or with the H3K4me3 and H4K16ac ChIP-seq data sets (Supplementary Data 2). We found robust upregulation of many of these genes in NAc of resilient mice (Fig. 3g).

Regulation of *Dicer1* and microRNA by β -catenin

One gene validated by qChIP and quantitative PCR (qPCR) was *Dicer1*, a critical component of microRNA (miRNA) biogenesis²³. Thus, selective enrichment of β -catenin binding at *Dicer1* in resilient mice (Fig. 4a),

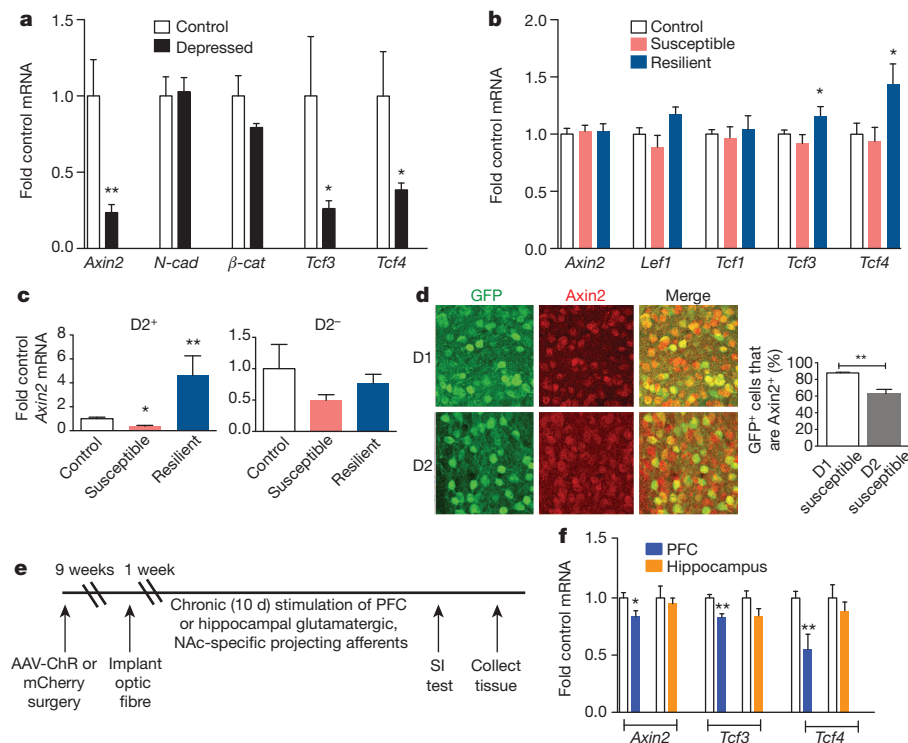


Figure 2 | Regulation of β -catenin signalling in human depression and mouse CSDS. **a**, mRNA from human NAc (*Axin2*: $**P < 0.01$; *Tcf3*: $*P < 0.05$; *Tcf4*: $*P < 0.05$, two-tailed *t*-test, $n = 6$ control, $n = 10$ depressed). **b**, mRNA from mouse control, susceptible, and resilient NAc 48 h post CSDS (*Tcf3*: $*P < 0.05$; *Tcf4*: $*P < 0.05$, one-way ANOVA, $n = 16$ control, $n = 12$ susceptible, $n = 9$ resilient). **c**, *Axin2* is upregulated in $D2^+$ MSNs only in resilience (*Axin2* $D2^+$: $**P < 0.01$, control versus resilient $P < 0.05$, $*P < 0.01$ susceptible versus resilient, $n = 4$ control, $n = 5$ susceptible, $n = 3$ resilient; *Axin2* $D2^-$: not significant, $P > 0.05$, $n = 3$ control, $n = 5$ susceptible, $n = 3$ resilient, one-way ANOVA). **d**, Percentage of cells positive for *Axin2* plus GFP in D1- or D2-GFP susceptible mice after CSDS ($**P < 0.01$, two-tailed *t*-test, $n = 3$ per group). **e**, Optogenetic stimulation protocol. **f**, mRNA expression in NAc after repeated stimulation from PFC or hippocampus in ChR2 versus mCherry (*Axin2*: $*P < 0.05$, $n = 6$ mCherry, $n = 5$ ChR; *Tcf3*: $**P < 0.01$, $n = 6$ mCherry, $n = 4$ ChR; *Tcf4*: $**P < 0.01$, $n = 6$ mCherry, $n = 4$ ChR, two-tailed *t*-test). Human data are from one experiment, all other data are representative of at least two experiments. All data presented as mean and s.e.m.

and subsequent validation of this effect (Fig. 3f, g), indicated that *Dicer1* represents a robust target of β -catenin in NAc. To study the behavioural effects of *Dicer1*, we knocked it down locally in NAc (Extended Data Fig. 10), and conducted sub-threshold defeat. Control animals injected with HSV-GFP displayed normal social interaction; however, animals with *Dicer1* knockdown demonstrated social avoidance (Fig. 4b), which mimicked the effects of blocking β -catenin signalling (Fig. 1). Importantly, we can rule out confounding effects of long-term *Dicer1* loss on

neuronal viability²⁴, because our experimental paradigm was limited to two weeks.

To assess whether the behavioural effect of *Dicer1* was related to β -catenin signalling, we first expressed HSV-Cre or HSV-GFP in NAc of floxed *Dicer1* mice and found no difference in social interaction under baseline, non-stressed conditions (Fig. 4c). We then injected all mice with HSV- β -catenin in NAc and subjected them to ASD. β -catenin overexpression blocked the development of social avoidance in mice

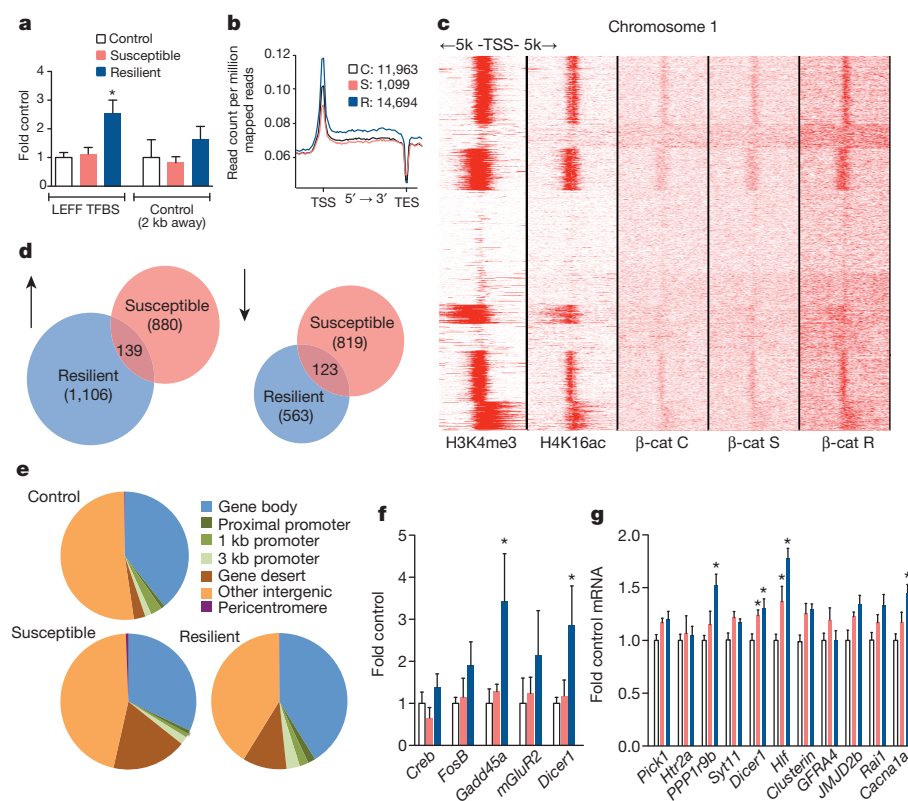


Figure 3 | β -catenin ChIP-seq in NAc 48 h post CSDS. **a**, qChIP validation of β -catenin ChIP ($*P < 0.05$, one-way ANOVA, post-hoc test control versus resilient and susceptible versus resilient at LEFF transcription factor binding site (TFBS) of a CaMKIV gene, $n = 4$ per group). **b**, Plot of β -catenin binding across genic regions. TSS, transcription start site; TES, transcription end site. Individual peak numbers per condition indicated in inset. **c**, Heat map showing β -catenin binding 5 kb up- and downstream of TSSs on chromosome 1 in control (C), susceptible (S), and resilient (R) NAc; binding profiles of H3K4me3 and H4K16ac under basal conditions are also shown. **d**, Number of increased (up arrow) versus decreased (down arrow) β -catenin binding sites at promoters in resilient versus control or susceptible versus control conditions. **e**, Genome-wide distribution of β -catenin binding. **f**, qChIP validation of ChIP-seq (*Gadd45a*: $*P < 0.05$, one-way ANOVA; *Dicer1*: $*P < 0.05$, one-way ANOVA, $n = 4$ control, susceptible, $n = 3$ resilient). **g**, mRNA validation of β -catenin ChIP-seq (*Dicer1*: $*P < 0.01$, one-way ANOVA, $n = 13$ control, $n = 11$ susceptible, $n = 7$ resilient). Data presented as mean and s.e.m. and are representative of at least two experiments. Colour-coding in **f** and **g** as in **a**.

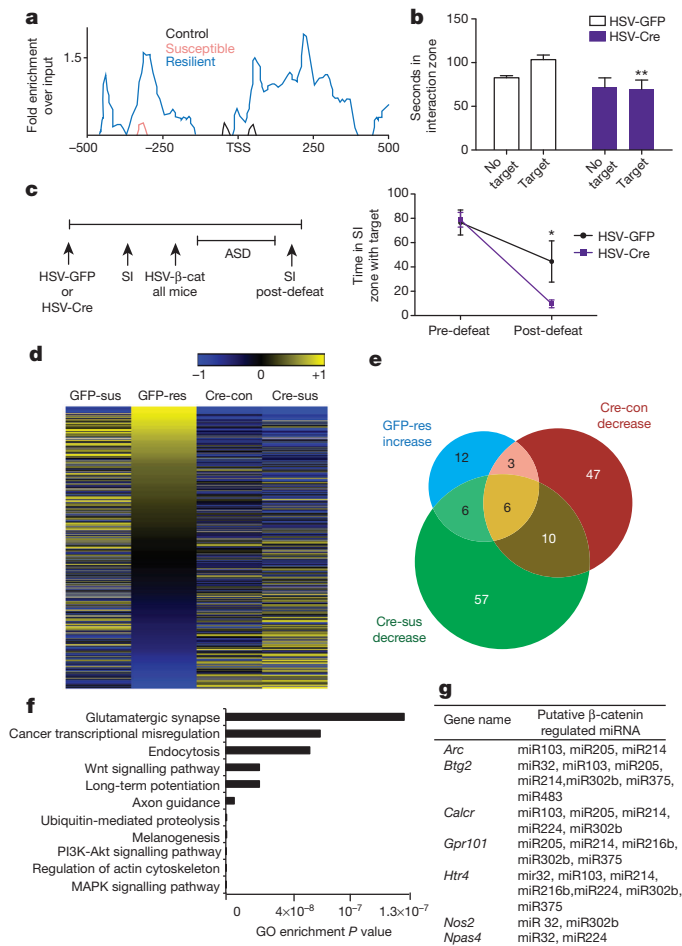


Figure 4 | Dicer1 bridges β-catenin and miRNA regulation in CSDS.

a, β-catenin ChIP-seq enrichment around the Dicer1 TSS. **b**, Effect of NAc Dicer1 knockdown (HSV-Cre) in sub-threshold defeat with HSV-GFP as control (** $P < 0.01$, effect of virus, two-way ANOVA, $n = 7$ Cre, $n = 8$ GFP). **c**, Left, schematic of floxed Dicer1 deletion followed by β-catenin rescue; right, social interaction (SI) before and after ASD with HSV-β-catenin (* $P < 0.05$, interaction effect, matching two-way ANOVA, $n = 7$ per group). **d**, Heat map of CSDS-regulated miRNA expression changes with (Cre) or without (GFP) β-catenin knockdown. Log₂-fold changes of all altered miRNAs among all groups are shown. **e**, Venn diagram showing increased miRNAs in GFP-resilient mice (GFP-res) overlap with decreased miRNAs in β-catenin knockout in non-stressed (Cre-con) or susceptible (Cre-sus) animals. **f**, Top 11 most enriched gene ontology terms of target genes of overlapping miRNAs in panel e. **g**, Predicted targets of β-catenin-dependent miRNAs display downregulation by mRNA-seq in resilient mice after CSDS. Data presented as mean and s.e.m. and are representative of at least two experiments.

expressing normal Dicer1 levels, but not in mice with NAc Dicer1 knockdown (Fig. 4c). This indicates that at least part of the pro-resilient effect of β-catenin is mediated through Dicer1.

Finally, these data prompted us to examine the global miRNA profile in NAc in response to CSDS and study its dependence on β-catenin. We injected an adeno-associated virus (AAV) vector expressing GFP or Cre in NAc of floxed β-catenin mice, subjected them to CSDS or control conditions, and performed small RNA sequencing (Supplementary Table 3). We first compared each group—GFP susceptible (GFP-sus), GFP resilient (GFP-res), Cre control (Cre-con), and Cre susceptible (Cre-sus)—to the ‘GFP-con’ condition. We could not study the Cre resilient condition, because virtually no mice are resilient upon β-catenin knockout from NAc. We found downregulation of numerous miRNAs, including many that were upregulated in resilience, when β-catenin was knocked out from control animals (Cre-con, Fig. 4d, Supplementary Table 4). Interestingly, a smaller subset of miRNAs was upregulated

following β-catenin knockout, which may represent miRNAs that are regulated by repressive factors under β-catenin control. We identified 66 miRNAs that were significantly downregulated in NAc after β-catenin deletion (Cre-con, Fig. 4e). We also identified downregulated miRNAs ($n = 79$) in the Cre-sus condition, many of which were decreased in Cre-con, further substantiating our hypothesis that pro-adaptive miRNA responses are lost in the absence of β-catenin, enhancing susceptibility to stress (Fig. 4e). miRNAs that overlapped between any two groups (up in GFP-res, but down in Cre-con or Cre-sus), presumably represent the most biologically important β-catenin- and stress-regulated miRNAs (Fig. 4e, Supplementary Table 5). This subset controls several meaningful gene categories (Fig. 4f), including Wnt and glutamatergic signalling. Finally, to identify potential miRNA targets, we overlapped predicted targets of these β-catenin-regulated miRNAs (Supplementary Table 5) with mRNA-seq data from NAc after CSDS. We thus found several interesting, novel genes to be significantly repressed in resilience (Fig. 4g).

We also examined other small RNAs for regulation by CSDS. Piwi-interacting RNAs (piRNAs), small RNAs widely studied in germ line cells, were detected recently in brain and found to play a functional role in spine morphology and synaptic plasticity^{25,26}. 163 piRNAs were detectable in our data set with read counts in at least one condition, supporting the notion of piRNA expression in brain (Supplementary Table 6). Although the majority of them were expressed at low levels, approximately 20 piRNAs appear to be regulated by CSDS (Supplementary Table 7). Examining additional small RNA categories that might be regulated by Dicer1 revealed several differentially expressed candidates (Supplementary Table 8), laying the groundwork for future investigation.

Discussion

The present study demonstrates that β-catenin in D2 MSNs activates a network in NAc that mediates behavioural resilience, whereas deficits in this pathway contribute to depression-related pathology. PFC inputs to NAc appear to be particularly important in controlling this β-catenin regulation. D2 MSNs, which comprise the indirect or ‘no-go’ pathway^{27–30}, may be more important for mediating flexible behavioural choices in aversive contexts compared to reward-motivated behaviour^{31–33}. We thus posit that enhanced β-catenin signalling in NAc D2 MSNs of resilient mice permits increased behavioural flexibility, which allows them, despite having the same experience as susceptible mice, to overcome generalizing avoidance of all mice, a process independent of hedonic responses. This has parallels in humans: resilient individuals are more successful at managing stress and recovering from it³⁴.

Our β-catenin ChIP-seq approach provides a valuable resource for mining the molecular targets that drive resilience. One validated target is Dicer1, which establishes a novel connection between β-catenin signalling and miRNAs in brain. Among the regulated miRNAs are those that feedback and regulate β-catenin signalling³⁵. The cell type-specific role of β-catenin, and the inherent complexity of stress susceptibility versus resilience, which involves many additional regulatory steps beyond Dicer1, presumably explains the relatively small number of β-catenin-dependent miRNAs observed in this study. miRNAs provide a crucial layer of post-transcriptional gene regulation in neural development, plasticity, and in an increasing number of brain disorders^{36–38}. The present study, by identifying specific miRNAs associated with stress susceptibility or resilience, offers a template for future studies to induce resilience in inherently more susceptible individuals.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 August 2013; accepted 20 October 2014.

Published online 12 November 2014.

- Madsen, T. M., Newton, S. S., Eaton, M. E., Russell, D. S. & Duman, R. S. Chronic electroconvulsive seizure up-regulates β-catenin expression in rat hippocampus: role in adult neurogenesis. *Biol. Psychiatry* **54**, 1006–1014 (2003).

2. Beaulieu, J.-M. *et al.* Lithium antagonizes dopamine-dependent behaviors mediated by an AKT/glycogen synthase kinase 3 signaling cascade. *Proc. Natl Acad. Sci. USA* **101**, 5099–5104 (2004).
3. Gould, T. D. *et al.* Beta-catenin overexpression in the mouse brain phenocopies lithium-sensitive behaviors. *Neuropsychopharmacology* **32**, 2173–2183 (2007).
4. Li, X. & Jope, R. S. Is glycogen synthase kinase-3 a central modulator in mood regulation? *Neuropsychopharmacology* **35**, 2143–2154 (2010).
5. Brennand, K. J. *et al.* Modelling schizophrenia using human induced pluripotent stem cells. *Nature* **473**, 221–225 (2011).
6. Behrens, J., von Kries, J., Kühl, M. & Brühn, L. Functional interaction of β -catenin with the transcription factor LEF-1. *Nature* **382**, 638–642 (1996).
7. Molenaar, M. *et al.* XTcf-3 transcription factor mediates β -catenin-induced axis formation in *Xenopus* embryos. *Cell* **86**, 391–399 (1996).
8. van de Wetering, M. *et al.* Armadillo coactivates transcription driven by the product of the *Drosophila* segment polarity gene dTCF. *Cell* **88**, 789–799 (1997).
9. Wilkinson, M. B. *et al.* A novel role of the WNT-dishevelled-GSK3 β signaling cascade in the mouse nucleus accumbens in a social defeat model of depression. *J. Neurosci.* **31**, 9084–9092 (2011).
10. Sadot, E. *et al.* Regulation of S33/S37 phosphorylated β -catenin in normal and transformed cells. *J. Cell Sci.* **115**, 2771–2780 (2002).
11. Berton, O. *et al.* Essential role of BDNF in the mesolimbic dopamine pathway in social defeat stress. *Science* **311**, 864–868 (2006).
12. Krishnan, V. *et al.* Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions. *Cell* **131**, 391–404 (2007).
13. Kolligs, F. T., Hu, G., Dang, C. V. & Fearon, E. R. Neoplastic transformation of RK3E by mutant β -catenin requires deregulation of Tcf/Lef transcription but not activation of *c-myc* expression. *Mol. Cell Biol.* **19**, 5696–5706 (1999).
14. Wang, Z. *et al.* β -catenin promotes survival of renal epithelial cells by inhibiting Bax. *J. Am. Soc. Nephrol.* **20**, 1919–1928 (2009).
15. Rada, P. *et al.* Glutamate release in the nucleus accumbens is involved in behavioral depression during the Porsolt swim test. *Neuroscience* **119**, 557–565 (2003).
16. Abe, K. & Takeichi, M. NMDA-receptor activation induces calpain-mediated β -catenin cleavages for triggering gene expression. *Neuron* **53**, 387–397 (2007).
17. Tye, K. M. *et al.* Amygdala circuitry mediating reversible and bidirectional control of anxiety. *Nature* **471**, 358–362 (2011).
18. Britt, J. P. *et al.* Synaptic and behavioral profile of multiple glutamatergic inputs to the nucleus accumbens. *Neuron* **76**, 790–803 (2012).
19. Feng, J. *et al.* Chronic cocaine-regulated epigenomic changes in mouse nucleus accumbens. *Genome Biol.* **15**, R65 (2014).
20. Shen, L. *et al.* diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. *PLoS ONE* **8**, e65598 (2013).
21. Hödar, C. *et al.* Genome-wide identification of new Wnt/ β -catenin target genes in the human genome using CART method. *BMC Genomics* **11**, 348 (2010).
22. Wexler, E. M. *et al.* Genome-wide analysis of a Wnt1-regulated transcriptional network implicates neurodegenerative pathways. *Sci. Signal.* **4**, ra65 (2011).
23. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
24. Cuellar, T. L. *et al.* Dicer loss in striatal neurons produces behavioral and neuroanatomical phenotypes in the absence of neurodegeneration. *Proc. Natl Acad. Sci. USA* **105**, 5614–5619 (2008).
25. Lee, E. J. *et al.* Identification of piRNAs in the central nervous system. *RNA* **17**, 1090–1099 (2011).
26. Rajasethupathy, P. *et al.* A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. *Cell* **149**, 693–707 (2012).
27. Graybiel, A. M. The basal ganglia. *Curr. Biol.* **10**, R509–R511 (2000).
28. Gerfen, C. R. The neostriatal mosaic: multiple levels of compartmental organization in the basal ganglia. *Annu. Rev. Neurosci.* **15**, 285–320 (1992).
29. Kravitz, A. V., Tye, L. D. & Kreitzer, A. C. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nature Neurosci.* **15**, 816–818 (2012).
30. Lobo, M. K. & Nestler, E. J. The striatal balancing act in drug addiction: distinct roles of direct and indirect pathway medium spiny neurons. *Front. Neuroanat.* **5**, 41 (2011).
31. Hikida, T., Kimura, K., Wada, N., Funabiki, K. & Nakanishi, S. Distinct roles of synaptic transmission in direct and indirect striatal pathways to reward and aversive behavior. *Neuron* **66**, 896–907 (2010).
32. Darvas, M. & Palmiter, R. Contributions of striatal dopamine signaling to the modulation of cognitive flexibility. *Biol. Psychiatry* **69**, 704–707 (2011).
33. Yawata, S., Yamaguchi, T., Danjo, T., Hikida, T. & Nakanishi, S. Pathway-specific control of reward learning and its flexibility via selective dopamine receptors in the nucleus accumbens. *Proc. Natl Acad. Sci. USA* **109**, 12764–12769 (2012).
34. Southwick, S. M. & Charney, D. S. The science of resilience: implications for the prevention and treatment of depression. *Science* **338**, 79–82 (2012).
35. Veronese, A. *et al.* Mutated β -catenin evades a microRNA-dependent regulatory loop. *Proc. Natl Acad. Sci. USA* **108**, 4840–4845 (2011).
36. Kosik, K. S. The neuronal microRNA system. *Nature Rev. Neurosci.* **7**, 911–920 (2006).
37. Im, H.-I. & Kenny, P. J. MicroRNAs in neuronal function and dysfunction. *Trends Neurosci.* **35**, 325–334 (2012).
38. Issler, O. *et al.* MicroRNA 135 is essential for chronic stress resiliency, antidepressant efficacy, and intact serotonergic activity. *Neuron* **83**, 344–360 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank O. Jabado and M. Mahajan for support and S. Borkan for providing β -catenin constructs. This work was supported by grants from the National Institute of Mental Health and the Hope for Depression Research Foundation (HDRF).

Author Contributions C.D. and J.F. conceived the project, designed research, conducted experiments, interpreted the results, and wrote the manuscript; H.S., M.S.M.-R., D.D.-W., K.S., R.B., B.L., E.R., P.K., V.V., D.F., C.P., E.C., J.K. and E.M. conducted experiments; S.G., C.T. provided reagents and tools; R.N. conducted experiments and provided reagents; N.S., X.L. performed bioinformatic analysis; L.S. performed and supervised bioinformatic analysis; E.J.N. conceived the project, designed and supervised research, interpreted the results, and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information All sequencing data have been deposited into the Gene Expression Omnibus with accession numbers GSE61294 and GSE61295. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.J.N. (eric.nestler@mssm.edu) or L.S. (li.shen@mssm.edu).

METHODS

Animals. For all experiments, 7–9-week-old male mice were used. Unless otherwise noted for transgenic lines, c57bl/6 mice from Jackson Laboratories were used. All mice were housed on a 12-h light/dark cycle with ad libitum access to food and water. CD1 retired breeder mice were obtained from Charles River Laboratories. The following transgenic mouse lines were used. From Jackson Laboratories: β -catenin conditional floxed mice (stock no. 004152) and Dicer1 conditional floxed mice (stock no. 006001). Additionally, D1-Cre, D2-Cre, D1-GFP, and D2-GFP male mice that were backcrossed to a c57bl/6 background were used for experiments as described in the text. For the D1-Cre/D2-Cre cell-type specific overexpression experiments, wild-type littermates were used as controls. The Mount Sinai Institutional Animal Care and Use Committee approved all animal protocols used in this study. For all experiments, extensive laboratory experience was used to estimate required sample sizes. Animals were randomly assigned to experimental groups and whenever possible, experimenters were blinded to the group. (For example, in behavioural experiments by assigning numbers to animals and in IHC experiments by hiding group designation until after quantification and analysis.)

Viral-mediated gene transfer. Stereotactic surgery was performed on mice under ketamine/xylazine anaesthesia. Vectors were infused bilaterally into NAc at a rate of $0.1 \mu\text{l min}^{-1}$ with the following coordinates: +1.6 mm anterior-posterior (A/P), +1.5 mm medial-lateral (M/L), ± 4.4 mm dorsal-ventral (D/V) from bregma. A total of $0.5 \mu\text{l}$ per side was infused except for the HSV-LS virus, in which case $0.7 \mu\text{l}$ was infused total. All vectors used were cloned into p1005 HSV or LS1L HSV. Mouse β -catenin constructs were provided by S. Borkan (Boston University). Wild-type and dominant negative constructs were used, with the dominant negative construct containing amino and carboxy-terminal truncations. Because this is a complicated mutant, we behaviourally validated it by demonstrating a failure to rescue β -catenin loss of function impairments in social interaction (Extended Data Fig. 2c). Human β -catenin S33Y construct (Addgene Plasmid no. 19286) was originally from E. Fearon (Michigan). This mutant contains an S33Y mutation that prevents phosphorylation at Ser 33 by GSK3 β , thus preventing β -catenin degradation. For cell-type-specific overexpression, an HSV carrying β -catenin in a lox-stop cassette was used (Supplementary Fig. 2a) in conjunction with D1- and D2-Cre transgenic mouse lines. Viral-Cre was used for local knockdown of β -catenin or Dicer1 in conditional floxed mice.

Behaviour. 10-day chronic social defeat stress (CSDS), an accelerated 4-day defeat procedure (ASD), and a sub-threshold defeat procedure have been described previously and represent an ethologically validated model of depression^{9,11,12}. We used ASD over 4 days (4 days of defeat, twice a day) to coincide with periods of maximal HSV-mediated transgene expression in some experiments as described, which induces the same degree of behavioural deficits in normal mice as our standard 10-day CSDS procedure. For all defeats, social interaction was measured either 24 h or 1 week following the last defeat. For all tissue analysis, including ChIP, mice were killed 48 h after the last defeat (24 h after social interaction) of a 10 day CSDS paradigm unless otherwise specified. Elevated plus maze and forced swim tests were performed as described previously¹².

Post-mortem human tissue. Human post-mortem NAc complementary DNA was generated and analysed as before⁹. Briefly, brain tissue was obtained from the Dallas Brain Collection, where tissue is collected from the Dallas Medical Examiner's Office and UT Southwestern's Tissue Transplant Program following consent of next-of-kin. Tissue was analysed and matched for age, post-mortem interval, RNA integrity number (RIN), and pH (see Supplementary Table 1) and this same tissue set was used in previously published work³⁹. Samples were subjected to a standard dissection before snap freezing in -40°C isopentane and storage at -80°C ; further dissection of NAc was performed on frozen tissue. The UT Southwestern Institutional Review Board reviewed and approved the collection of this tissue for research use. We should note that there was no difference in expression of Axin2 between medicated and unmedicated depressed patients, although all patients were clinically depressed at their time of death (Supplementary Fig. 4a). We thus combined the medicated and unmedicated groups into one overall depressed group as presented in Fig. 2.

RNA isolation and qPCR. RNA was extracted and purified using a protocol combining TRIzol/chloroform extraction with the Qiagen RNeasy Micro kit, with a motorized mini-pestle vibrator to homogenize the tissue. After extraction, purity and concentration were measured on a NanoDrop spectrophotometer. RNA was then reverse transcribed into cDNA with the iScript DNA synthesis kit (Bio-Rad). GAPDH was used to normalize quantification. Primers were designed to flank exon/intron boundaries and were created using the open-source software Primer3. Real-time qPCR analysis was performed with the $\Delta\Delta\text{C}_t$ method to obtain relative fold-change of expression as compared to control samples⁴⁰. BLAST and dissociation curve analysis was also performed to ensure specificity of primer design.

Western blotting. NAc was dissected bilaterally using 14 gauge steel circular punches. The tissue was then sonicated in radioimmunoprecipitation assay (RIPA) buffer

with a desktop sonicator (10 mM Tris, pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% Triton X-100, 1% sodium deoxycholate, with protease and phosphatase inhibitors) and centrifuged. The supernatant was collected and the protein concentration was quantified using the Lowry method. Laemmli sample buffer was added to the protein lysate and equal amounts of protein were loaded onto precast SDS-PAGE gels with molecular weight ladders. Samples were transferred to activated PVDF membranes, blocked, and incubated in primary antibody overnight. Blots were washed, and then incubated with Licor secondary fluorescent antibodies. After further washing, the blots were scanned and images analysed with ImageJ software. The following antibodies were used: phospho-Ser 675 β -catenin (Cell Signaling no. 4176; Ser 675 is phosphorylated by PKA), total β -catenin (Cell Signaling no. 9562), GAPDH, β -tubulin, and total H3. All antibodies are commercially available and have been validated for use in the laboratory. Pre-incubating the tissue with calf intestinal phosphatase and demonstrating a decrease in signal was performed to validate the phospho-Ser 675 β -catenin antibody.

Optogenetics. For glutamatergic nerve terminal stimulation, mice were injected unilaterally with AAV-CAMKIIa-ChR2-mCherry or AAV-CAMKIIa-mCherry with the coordinates of: (−3.6 A/P, +3.05 M/L, −4.85 D/V) for ventral hippocampus and (+1.9 A/P, +0.5 M/L, −3.0 D/V) for PFC unilaterally. After 9 weeks of recovery to allow for expression in terminals, a second stereotactic surgery was performed to implant an optic fibre targeting the NAc shell with coordinates of (+1.4 A/P, +1.5 M/L, −4 D/V), again unilaterally, ipsilateral to virus expression. After allowing one week for recovery, the mice underwent 10 days of daily 5-min stimulation sessions outside of their home cage as described^{41,42}. Stimulation parameters were either 20 Hz, 30 pulses per burst, with 10 s between bursts (hippocampus); or 30 Hz, 90 pulses per burst, 10 s between bursts (PFC) to roughly balance the relative intensity of NAc innervation from these two afferent regions. Unilateral NAc tissue was then dissected 48 h later for biochemical experiments. Constructs and stimulation parameters have been previously validated^{17,18}. AAV-ChR2 was used to stimulate VTA cell bodies with a phasic protocol (20 Hz, 5 spikes per burst, 10 s between bursts) given susceptible mice exhibit increased firing rate and bursting events following defeat^{12,43}.

Co-immunoprecipitation (Co-IP). A co-IP kit (Roche) was used as follows. 4 punches of NAc were lysed in 300 μl of the provided lysis buffer. 10% total lysate was reserved and the rest was centrifuged and the supernatant transferred to a clean microcentrifuge tube. It was pre-cleared by incubation with protein G-agarose for 3 h on a rotator at 4°C . The beads were centrifuged, and the supernatant was transferred to fresh tubes, where they were incubated with 5 μl of β -catenin antibody (Cell Signaling no. 9581) for one hour before 50 μl of a homogeneous protein G-agarose suspension was added and then incubated overnight at 4°C on a rotator. The complexes were centrifuged and the supernatant was removed, the beads were washed twice with lysis buffer 1, twice with buffer 2, and once with buffer 3. Protein sample buffer was added and the samples boiled for 3 min. Complexes were then analysed as described under western blotting.

Nuclear/cytoplasmic fractionation. NAc punches were homogenized with a glass Dounce tissue grinder and loose pestle in Buffer A (1 M Tris-HCl, 1 M sucrose, 1 M DTT, protease and phosphatase inhibitors). 10% of the lysate was reserved to assay total protein levels, and the rest was centrifuged at 1,450 r.c.f. in an Eppendorf 5417c centrifuge for 10 min. The supernatants were then removed, centrifuged at 5,970 r.c.f. in the same centrifuge for 7 min, and the resulting supernatants were stocked as the cytoplasmic fraction. Buffer B (1 M Tris-HCl pH 7.5, 0.1 M EDTA, 0.1 M EGTA, 1 M sucrose, 1 M DTT, 10% NP-40, protease and phosphatase inhibitors) was added to the pellets from the first centrifugation and the samples were kept on ice for 10 min before centrifuging again at 1,450 r.c.f. in the same centrifuge for 10 min. The supernatants were discarded and Buffer C (1 M Tris-HCl, 37.5% glycerol, 5 M NaCl, 0.1 M EDTA, 0.1 M EGTA, 1 M DTT, 10% NP-40, protease and phosphatase inhibitors) was used to re-suspend the nuclear fraction. The fractions were then processed for western blotting as above or further separated into chromatin and non-chromatin nuclear fractions. Tubulin and total H3 were used as loading controls and to verify appropriate cytoplasmic and nuclear enrichment.

Immunohistochemistry (IHC). Mice were anaesthetized with chloral hydrate followed by trans-cardial perfusion of 10 ml of filtered PBS, followed by 25 ml of filtered 4% paraformaldehyde (PFA) in PBS, pH 7.4. Brains were dissected out and post-fixed overnight in PFA. They were then rinsed in PBS and placed in 30% sucrose in PBS. For the IHC in Fig. 2, once the brains sank, coronal 35- μm sections through the NAc were taken on a freezing microtome and kept in PBS with 0.01% sodium-azide. The slices were washed $3\times$ in PBS for 10 min and then blocked for 3–4 h (3% normal goat serum, 0.3% TritonPBS) in net wells. They were incubated in primary antibody overnight diluted in block (rabbit Anti-Axin2, Abcam; Mouse Anti-GFP, Life Technologies) at 4 degrees. The slices were then washed $3\times$ in PBS, followed by a 1-h incubation in secondary antibody (Alexa-Fluor Anti-Rabbit & Anti-Mouse 680 & 800 diluted 1:1,000 in PBS). The slices were washed $4\times$ in PBS and then mounted on charged slides and allowed to dry overnight. They were

dehydrated, coverslipped with Depex mounting medium, and sealed with clear nail polish. Z-stacks were taken on a Zeiss LSM 710 confocal microscope at $\times 64$ magnification. Settings were kept identical for all images taken. The specificity of the Axin2 antibody was validated by competing the antibody with the immunizing protein. Average values of 3–5 images per mouse were used. For quantification purposes, the percent of Axin2⁺ GFP⁺ cells was counted per image, with Axin2⁺ being defined as $>20\%$ above background levels.

For the IHC in Supplementary Fig. 1, coronal sections (50 μ m thick) were made with a vibratome; sections were collected into antifreeze solution consisting of ethylene glycol, glycerol and PBS. Free-floating sections were blocked using 3% BSA in 0.1% PBST for 1 h. The sections were stained for 48 h at room temperature with primary antibody, and overnight with secondary antibody. The sections were mounted with Prolong Antifade reagent with DAPI (Life Technologies). Z-stacked images were acquired with a Zeiss LSM780 multi photon confocal system and processed using ImageJ. The number of GFP⁺ cells containing β -catenin staining was quantified by requiring the presence of β -catenin in the nucleus. To quantify β -catenin protein expression, we used the rabbit-conjugated primary antibody for total β -catenin (9562; Cell Signaling). We also amplified GFP staining using a chicken-conjugated primary antibody for GFP (Aves Laboratory). Stains were visualized using Chicken-Cy2 and Rabbit-Cy3 secondary antibodies (Jackson Immunolabs). **FACS.** D2⁺ and D2⁻ cells from the NAc of D2-GFP mice were isolated using a fluorescence-assisted cell sorting (FACS) protocol. Briefly, 48 h after our standard CSDS protocol, bilateral 12 gauge punches were taken from the NAc and digested with an enzyme cocktail for 30 min at 37 °C before being triturated to obtain a homogeneous cellular preparation. Cells were then processed through a gradient, washed, and labelled with DAPI (viability marker) before being processed through an Influx sorter (BD Bioscience). D2⁺ MSNs were sorted based on the size, internal complexity, and intensity of fluorescence with D2 cells emitting in the green channel (GFP). RNA was isolated using the Direct-zol RNA miniprep (Zymo Research) kit and cDNA was synthesized using the Iscript kit (Biorad). We confirmed the enrichment of D2 MSN-enriched genes in D2⁺ cells and D1 MSN-enriched genes in D2⁻ cells.

Quantitative chromatin immunoprecipitation (qChIP). Four 14 gauge NAc punches from each mouse were placed in 1% formaldehyde in 1 \times PBS to fix the DNA with the associated proteins. After 12 min on the rotator, 2 M glycine was added to stop the fixation for 5 min. The punches were then placed on ice and rinsed 5 \times with ice-cold PBS. Tissue from 5 animals were pooled at this point and homogenized in SDS lysis buffer (10% SDS, 1 M Tris-HCl, 0.5 M EDTA) with a desktop sonicator. ChIP dilution buffer (10% Triton X-100, 5 M NaCl, 1 M Tris-HCl pH 8.1, 0.5 M EDTA, 10% SDS and protease inhibitors) was added and the chromatin underwent high power sonication with the Bioruptor for 30 cycles of 30 s on/30 s off on high power. Conjugated magnetic beads were used to IP β -catenin with the ChIP-validated β -catenin antibody (Cell Signaling no. 4176) overnight in block solution (0.5% BSA in 1 \times PBS). The IP reaction was collected with a magnetic rack, washed, and both the input chromatin and the immunoprecipitated DNA were reverse cross-linked at 65 °C overnight. The DNA was then purified with RNase, proteinase K and the Qiagen PCR purification kit. The Qubit was used to quantify both the input and immunoprecipitated DNA and RT-PCR was used to quantify differential binding on the genomic DNA.

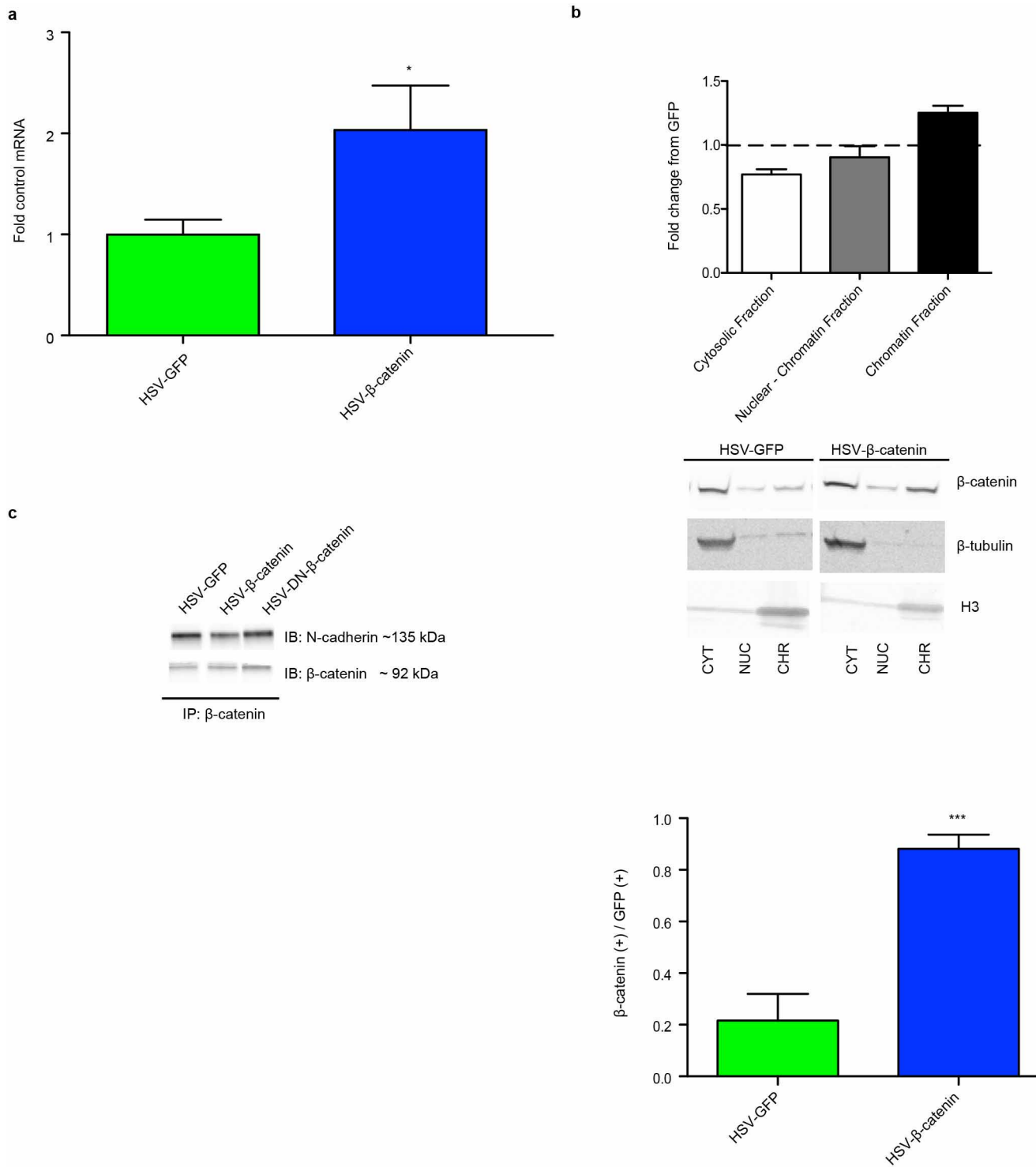
ChIP-seq. ChIP was performed for β -catenin as above. At the PCR purification step, however, 3 replicates were pooled onto one spin column so that each replicate became the pooled sample of 15 mice or 60 14 gauge NAc punches (4 NAc punches per animal), totalling ~ 100 mg wet weight tissue per library, following established protocols for brain. Animals were pooled so that social interaction times of different replicates across a group were approximately equal. ChIP-seq libraries were then prepared with the Illumina ChIP-seq kit as per their protocol. 2 replicates per condition were used for β -catenin while 3 replicates were used for histone mark experiments. Histone mark ChIP-seq was performed similarly except no further pooling was performed at the PCR purification step. Additionally, fragments from ~ 200 –400 bp were size-selected for sequencing for β -catenin to compensate for the decreased yield of DNA. Libraries were validated on the Bioanalyzer for appropriate size selection and amplification before being sent to the Mount Sinai Genomics Core for sequencing. Homer⁴⁴ was used to identify peaks in individual conditions and NGS plot was used to create genome-wide overviews of binding at gene bodies⁴⁵. We used hierarchical clustering based on the H3K4me3 data set to generate the heat map in Fig. 3c. To further validate our β -catenin data sets, we

found that, in all 3 treatment conditions, IPA pathway analysis independently identified β -catenin as an upstream regulator due to the enrichment of known β -catenin target genes.

Small RNA-seq and analysis. Small RNA (<200 bp) was isolated and enriched with Qiagen RNeasy mini kit (catalogue no. 74104) following instructions. The small RNA was then used for library preparation following Epicentre Scriptminer small RNA library kit (catalogue no. SMSPI0908) with optimization. In brief, a 3' adaptor tag was ligated to the small RNA, then a 5' adaptor oligonucleotide was attached following removal of excess 3' adaptor oligonucleotide with degradase. The Di-tagged RNA was purified with Zymo RNA Clean & Concentrator Kits (catalogue no. R1015) and followed with reverse transcription into cDNA using the cDNA Synthesis Primer and MMLV Reverse Transcriptase. After removing RNA template by addition of RNase, the di-tagged cDNA was amplified and individually barcoded with nine PCR cycles using indices and PCR primers provided in the kit. The library was purified with Zymo DNA Clean & Concentrator kit (catalogue no. D4003) and size selected with Pippin (Sage Science). The library concentration was confirmed on Agilent Bioanalyzer before sequencing. Multiplexed libraries were then pooled and sequenced on an Illumina HiSeq sequencer. In total, 4–12 libraries/condition were included in this study. Raw sequencing reads were processed by cutadapt (<https://code.google.com/p/cutadapt/>) to remove adaptor sequence at 3' end, and sequences shorter than 16 nucleotides after this were discarded. FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was applied to inspect the sequencing quality. We ensured our small RNA sequencing was of good quality as the majority of reads aligned to mature miRNAs (Supplementary Table 3). miRanalyzer was used to align the short reads to genomic annotations and quantify the expression of the non-coding RNAs⁴⁶. All miRNA annotations were downloaded from miRBase (v. 20)⁴⁷. piRNA annotations were merged from piRNA Bank and NCBI^{48,49}. tRNA and mRNA (RefSeq) annotations were downloaded from UCSC genome browser. The general ncRNA annotations were obtained from RFam (<http://rfam.xfam.org/>)⁵⁰. The pipeline was organized by Ruffus (<https://code.google.com/p/ruffus/>), and the code is accessible from GitHub (https://github.com/shenlab-sinai/miRNA_pipeline_for_miRanalyzer)⁵¹. The differential expression detection was applied by DESeq2 (<http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html>, <http://biorxiv.org/content/early/2014/02/19/002832>) with cut-offs of fold change 1.3 and P value < 0.05 .

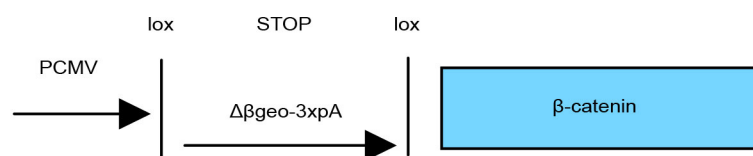
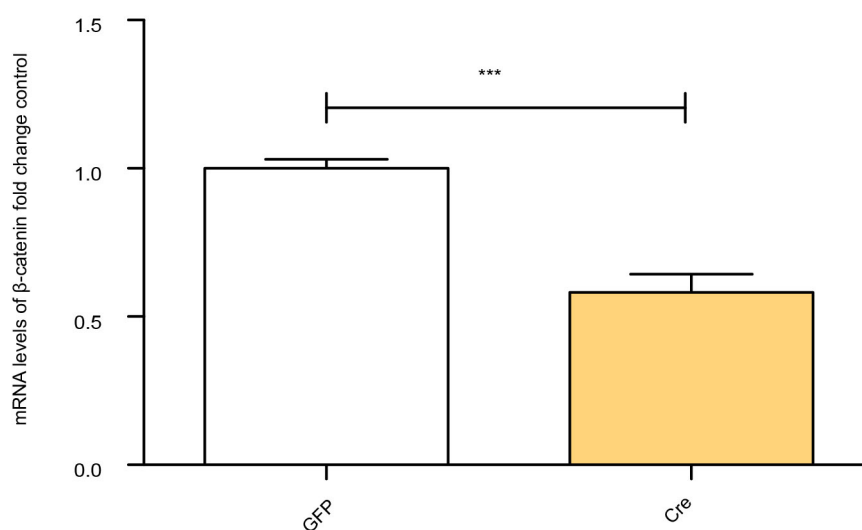
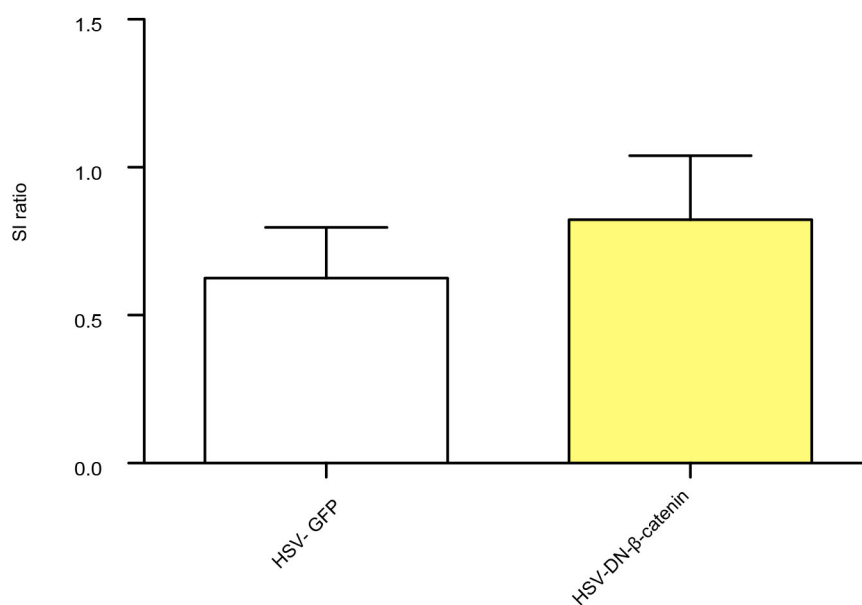
Statistics. One- or two-way analysis of variance (ANOVA) followed by Tukey's multiple comparison test, or two-way student's t -test were used for statistical analyses. All experiments represent at least 2–3 biological replicates unless otherwise indicated.

39. Robison, A. J. *et al.* Fluoxetine epigenetically alters the CaMKII α promoter in nucleus accumbens to regulate Δ FosB binding and antidepressant effects. *Neuropsychopharmacology* **39**, 1178–1186 (2014).
40. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C_T method. *Nature Protocols* **3**, 1101–1108 (2008).
41. Lobo, M. K. *et al.* Cell type-specific loss of BDNF signaling mimics optogenetic control of cocaine reward. *Science* **330**, 385–390 (2010).
42. Koo, J. W. *et al.* BDNF is a negative modulator of morphine action. *Science* **338**, 124–128 (2012).
43. Chaudhury, D. *et al.* Rapid regulation of depression-related behaviours by control of midbrain dopamine neurons. *Nature* **493**, 532–536 (2013).
44. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
45. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284 (2014).
46. Hackenberg, M., Rodríguez-Ezpeleta, N. & Aransay, A. M. MiRanalyzer: An update on the detection and analysis of microRNAs in high-throughput sequencing experiments. *Nucleic Acids Res.* **39** (Suppl. 2), W132–W138 (2011).
47. Griffiths-Jones, S. miRBase: the microRNA sequence database. *Methods Mol. Biol.* **342**, 129–138 (2006).
48. Sai Lakshmi, S. & Agrawal, S. piRNA Bank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.* **36** (Suppl. 1), W173–W177 (2008).
49. Kozomara, A. & Griffiths-Jones, S. MiRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68–D73 (2014).
50. Burge, S. W. *et al.* Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* **41**, D226–D232 (2013).
51. Goodstadt, L. Ruffus: A lightweight python library for computational pipelines. *Bioinformatics* **26**, 2778–2779 (2010).



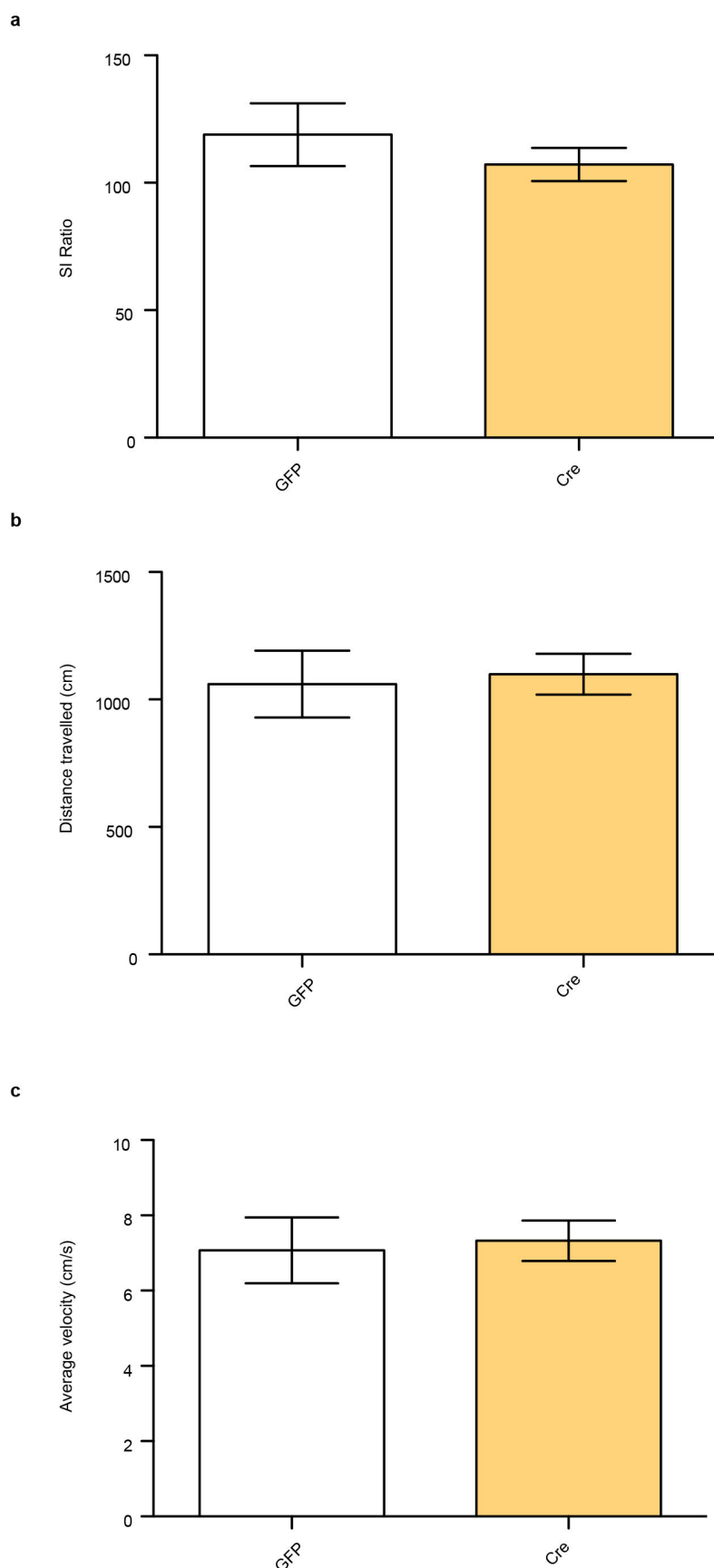
Extended Data Figure 1 | Validation of HSV-β-catenin. **a**, β-catenin mRNA levels following HSV-β-catenin versus HSV-GFP injection into NAc ($*P < 0.05$, two-tailed *t*-test, $n = 3$ per group). **b**, Top panel, subcellular fractionation of NAc lysates from HSV-GFP or HSV-β-catenin injected mice. Middle panel, representative western blots of data shown in panel **a**. CYT, cytosolic fraction; NUC, nuclear fraction (–chromatin); CHR, chromatin

fraction. Bottom panel, IHC of nuclear β-catenin 5 days post-injection with HSV-β-catenin versus HSV-GFP ($***P < 0.001$, two-tailed *t*-test, $n = 3$ per group). **c**, β-catenin IP on virus-injected NAc. IP results are representative of 5 replications. All other data shown are representative of at least two experiments. Data are presented as mean and s.e.m.

a**b****c**

Extended Data Figure 2 | Other $\beta\text{-catenin}$ manipulations. **a**, Schematic of Cre-dependent HSV-lox-stop (LSIL)- $\beta\text{-catenin}$ cassette. **b**, Validation of $\beta\text{-catenin}$ knockdown in the NAc of floxed $\beta\text{-catenin}$ mice (*** $P < 0.001$, two-tailed t -test, $n = 4$ GFP, $n = 5$ Cre). **c**, Failure of dominant negative $\beta\text{-catenin}$ to

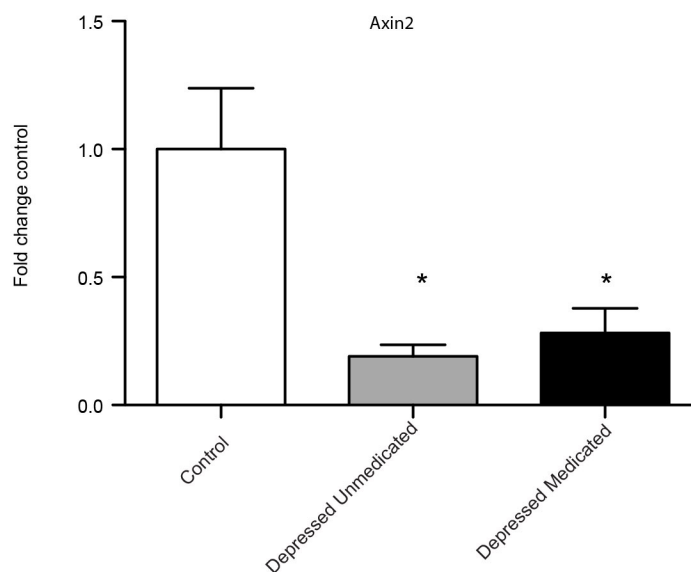
rescue social interaction as compared to GFP after previous excision of $\beta\text{-catenin}$ from NAc in floxed $\beta\text{-catenin}$ mice undergoing defeat ($P > 0.05$, two-tailed t -test, $n = 7$ per group). Data are presented as mean and s.e.m. All data shown are representative of at least two experiments.



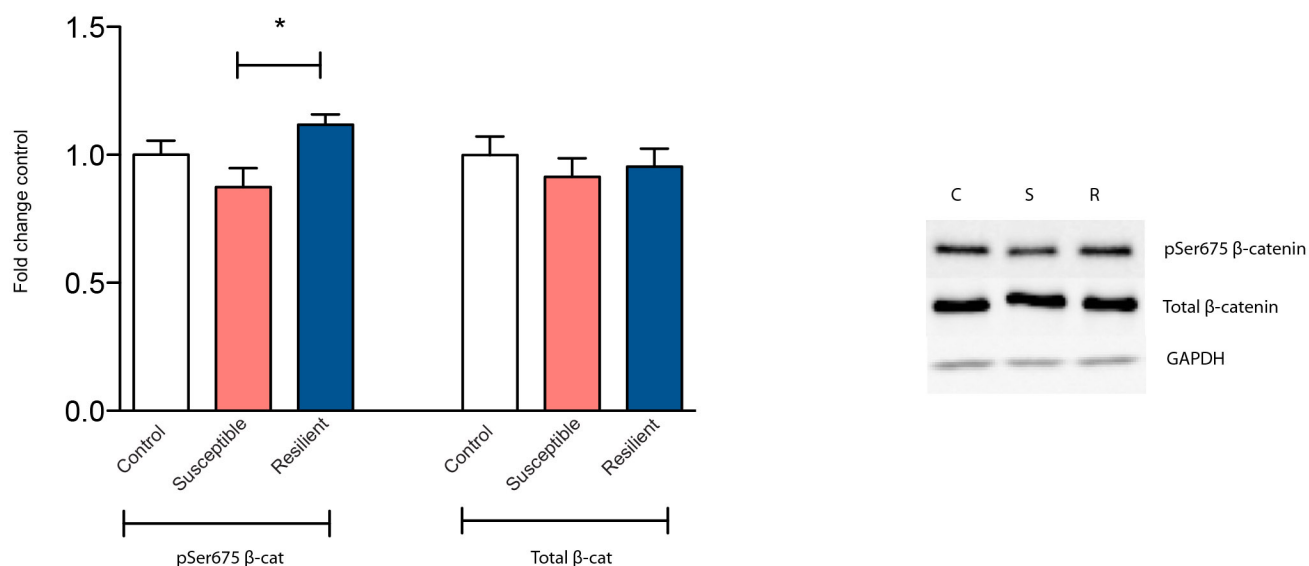
Extended Data Figure 3 | No effect of β -catenin deletion on baseline behaviours. **a**, Social interaction (SI) in control, non-stressed animals ($P > 0.05$, two-tailed t -test, $n = 5$ per group). **b**, Total distance travelled in

arena ($P > 0.05$, two-tailed t -test, $n = 5$ per group). **c**, Average velocity ($P > 0.05$, two-tailed t -test, $n = 5$ per group). Data are presented as mean and s.e.m. All data shown are representative of at least two experiments.

a

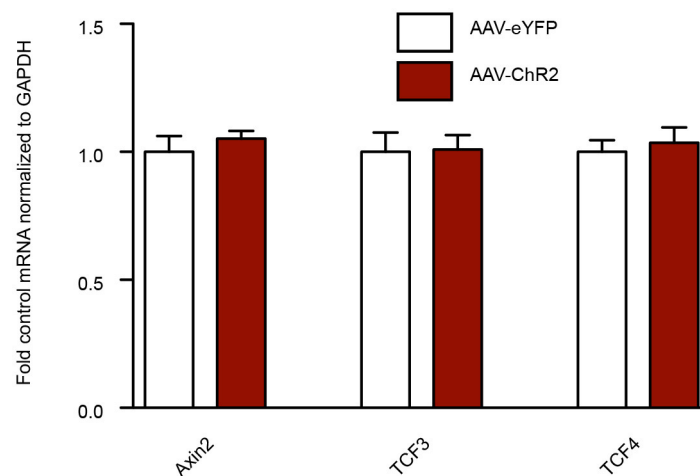


b



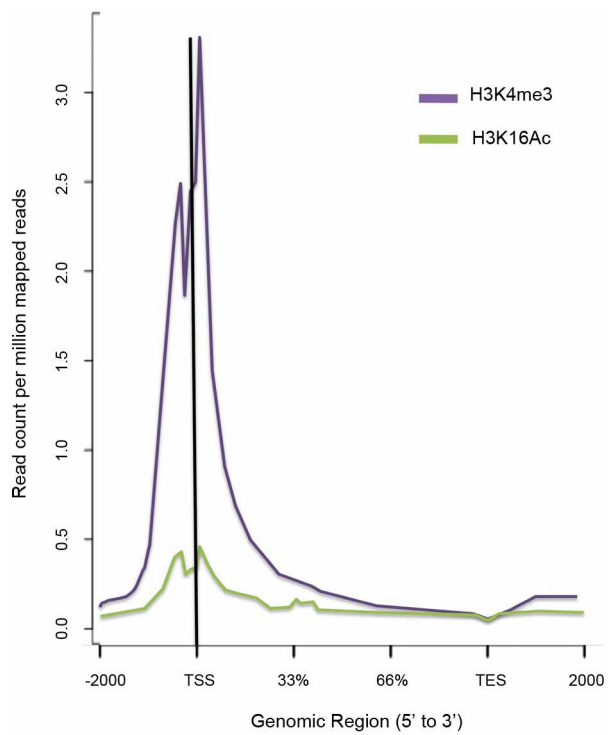
Extended Data Figure 4 | Regulation of β -catenin signalling in human depression and after CSDS in mice. **a**, Axin2 expression is suppressed in both medicated and unmedicated depressed patients, both groups of which were clinically depressed at their time of death ($P < 0.01$ one-way ANOVA, post-hoc test $P > 0.05$ between depressed unmedicated and medicated groups, $*P < 0.01$ for either depressed group versus control, $n = 6$ control, $n = 5$ unmedicated

depressed, medicated depressed). **b**, Phospho-Ser 675 β -catenin and total β -catenin levels from mouse control, susceptible, and resilient NAc 48 h post CSDS (phospho-Ser 675: $*P < 0.05$, one-way ANOVA, post-hoc test susceptible versus resilient, $n = 5$ for control, susceptible, $n = 8$ for resilient). Data are presented as mean and s.e.m. Human data are from one experiment. All other data shown are representative of two experiments.

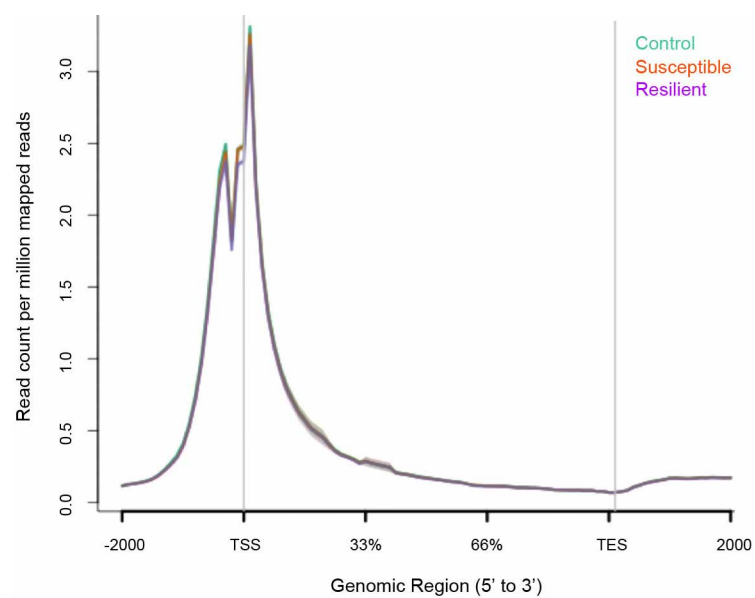


Extended Data Figure 5 | Repeated optogenetic burst stimulation of VTA cell bodies has no effect on canonical β -catenin signalling in NAc.
Experiment was performed as in Fig. 2 with the exception of the optic fibre,

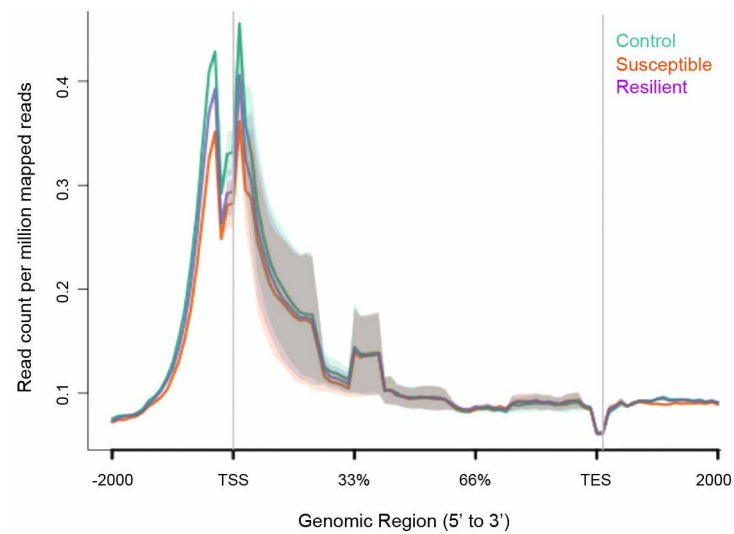
which was placed above VTA for cell body stimulation ($P > 0.05$, two-tailed t -test, $n = 8$ per group). Data are presented as mean and s.e.m. Data are from one experiment.



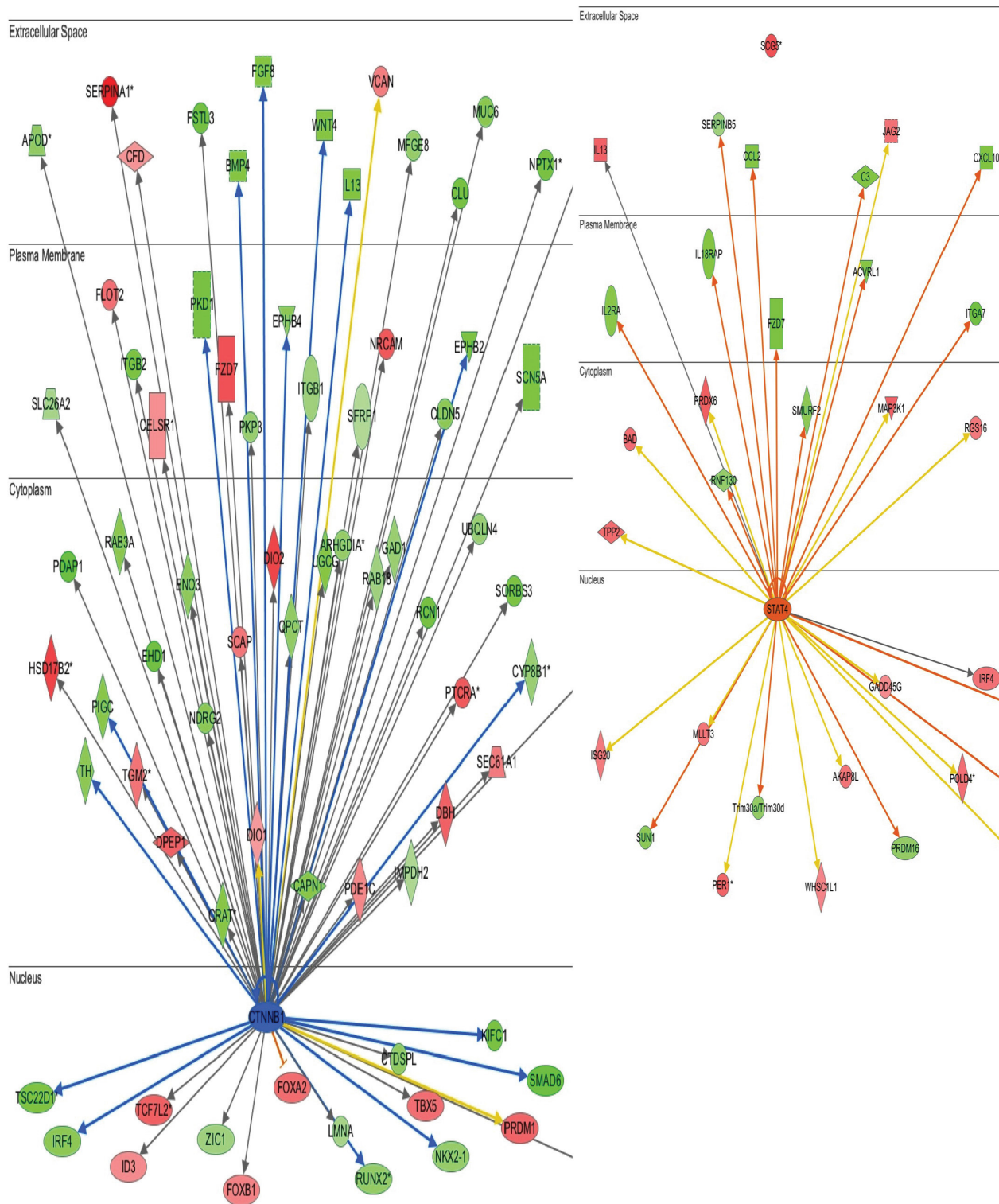
Extended Data Figure 6 | Genome-wide enrichment of H3K4me3 and H3K16Ac binding in NAc at TSSs. NGS plot was used to visualize binding patterns.



Extended Data Figure 7 | Genome-wide pattern of H3K4me3 binding to genic regions in NAc under control, susceptible (defeat), and resilient mice. Note the lack of difference across the three conditions. Data are from one experiment.



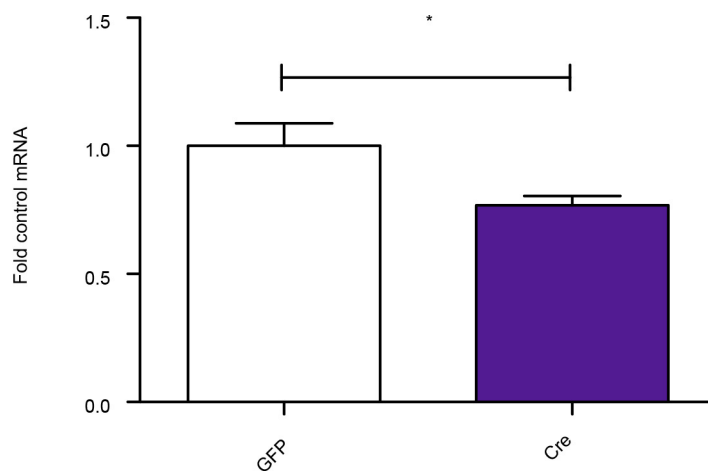
Extended Data Figure 8 | Genome-wide pattern of H4K16ac binding to genic regions in NAc under control, susceptible (defeat), and resilient mice. Note the lack of difference across the three conditions. Shading represents standard error. Data are from one experiment.



©2000-2012 Ingenuity Systems, Inc. All rights reserved.

Extended Data Figure 9 | Ingenuity pathway analysis (IPA) identifies a network of genes that show upregulated β -catenin binding at promoter regions in the NAc of resilient versus susceptible mice. Nodes represent differentially regulated genes, with green meaning up in resilient versus susceptible and red meaning down in resilient versus susceptible. The blue arrows indicate that the direction of regulation is consistent with IPA prediction of an upregulated β -catenin network in resilience; for example, a blue arrow means that a target gene that would be expected to be upregulated by

β -catenin is in fact upregulated in this list. In contrast, yellow arrows indicate that the regulation observed is inconsistent with expectations, while grey arrows indicate lack of pre-existing data to formulate expectations of β -catenin action. Left panel shows mostly expected regulation of the β -catenin network (that is, upregulation) in resilience; right panel shows non-specific changes occurring in a randomly chosen signal transducer and activator of transcription-4 (STAT4) network.



Extended Data Figure 10 | Validation of local Dicer1 knockdown. Note significant knockdown of Dicer expression in NAc after intra-NAc delivery of viral-Cre to floxed Dicer mice ($*P < 0.05$, two-tailed t -test, $n = 7$ GFP, $n = 6$ Cre). Data are presented as mean and s.e.m. and are representative of two experiments.

Deconstructing transcriptional heterogeneity in pluripotent stem cells

Roshan M. Kumar^{1,2*}, Patrick Cahan^{3*}, Alex K. Shalek⁴, Rahul Satija⁵, AJay Daley¹, Keyser¹, Hu Li⁶, Jin Zhang³, Keith Pardee^{1,2}, David Gennert⁵, John J. Trombetta⁵, Thomas C. Ferrante¹, Aviv Regev^{5,7}, George Q. Daley³ & James J. Collins^{1,2}

Pluripotent stem cells (PSCs) are capable of dynamic interconversion between distinct substates; however, the regulatory circuits specifying these states and enabling transitions between them are not well understood. Here we set out to characterize transcriptional heterogeneity in mouse PSCs by single-cell expression profiling under different chemical and genetic perturbations. Signalling factors and developmental regulators show highly variable expression, with expression states for some variable genes heritable through multiple cell divisions. Expression variability and population heterogeneity can be influenced by perturbation of signalling pathways and chromatin regulators. Notably, either removal of mature microRNAs or pharmacological blockage of signalling pathways drives PSCs into a low-noise ground state characterized by a reconfigured pluripotency network, enhanced self-renewal and a distinct chromatin state, an effect mediated by opposing microRNA families acting on the *Myc/Lin28/let-7* axis. These data provide insight into the nature of transcriptional heterogeneity in PSCs.

PSCs are defined by their unique capacity to differentiate into all the cell types of an organism, while self-renewing in culture. How PSCs reconcile pluripotency and self-renewal and decide among fate choices is a topic of intense interest, with relevance for regenerative medicine and developmental biology. Genomic maps of the regulatory circuitry underlying pluripotency reveal a network of sequence-specific autoregulatory transcription factors targeting self-renewal genes that are active in PSCs, as well as repressed lineage-specific developmental regulators that exist in a poised state and are capable of driving cells towards differentiated fates^{1–5}. These core transcription factors are thought to interact with chromatin modifiers, non-coding RNAs and external signalling pathways to maintain pluripotency. This self-sustaining transcriptional program becomes reactivated during reprogramming of somatic cells to pluripotency⁵.

The discoveries that levels of *Nanog* and other key PSC regulators fluctuate over time, that PSCs exist in multiple interconvertible states, and that distinct subpopulations of PSCs vary in their capacity to self-renew or differentiate, hint at the dynamism of the PSC transcriptional program^{6–13}, which may be fundamental to pluripotency^{14–23}. Here, we apply single-cell analytics to PSCs subjected to a range of perturbations to systematically dissect the factors underlying PSC heterogeneity. By doing so, we map the structure of gene expression variability in PSCs and identify regulatory circuits governing transitions between pluripotent cell states.

The landscape of gene expression variability in PSCs

To gain insight into the distinct substates of pluripotency, we first used single-cell RNA-seq^{24,25} to characterize the transcriptome of 183 individual mouse embryonic stem cells (mESCs) grown under standard culture conditions, in the presence of serum and leukaemia inhibitory

factor (LIF) (Extended Data Fig. 1, Supplementary Information and Supplementary Tables 1–3). Most cells (~92%) grouped together by principal component and cluster analysis, while 14 cells (8%) were characterized by reduced expression of fluctuating pluripotency regulators, which may indicate a distinct poised state (Extended Data Fig. 2 and Supplementary Information).

Some transcripts were detected in the vast majority of cells examined and showed a log-normal distribution of transcript abundance within the population, as for the core pluripotency regulator *Oct4* (also called *Pou5f1*) (Fig. 1a). Other transcripts had bimodal expression, present in some cells and absent in others, as for the pluripotency regulator *Esrrb* (Fig. 1a). Another set of genes displayed sporadic expression—that is, undetected in most cells but exhibiting relatively high expression in several cells—as for the lineage regulator and Polycomb target gene *Neurod1* (Fig. 1a, Extended Data Figs 3 and 4 and Supplementary Information). Expression distributions for 18 transcripts were validated by single-molecule fluorescence *in situ* hybridization (smFISH; Fig. 1b and Extended Data Fig. 3). Stable and variable regulators were also identified in neural precursor cells derived from mESCs (Supplementary Information, Supplementary Table 4 and Extended Data Fig. 4), suggesting that this may be a general feature of progenitor cell regulatory networks.

We examined functional gene categories for gene sets that showed more uniform or noisier expression in PSCs as compared to control gene sets. Genes involved in housekeeping and metabolic functions displayed relatively uniform expression, while previously identified targets²⁶ of the Polycomb family of epigenetic regulators in PSCs exhibited greater variability (Fig. 1c and Extended Data Fig. 4). The Polycomb target genes include many developmental regulators and signalling factors governing lineage specification, and are thought to exist in a repressed yet poised state with a unique chromatin signature in PSCs^{26,27}.

¹Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02115, USA. ²Howard Hughes Medical Institute, Department of Biomedical Engineering, Center of Synthetic Biology, Boston University, Boston, Massachusetts 02215, USA. ³Stem Cell Transplantation Program, Division of Pediatric Hematology and Oncology, Manton Center for Orphan Disease Research, Howard Hughes Medical Institute, Boston Children's Hospital and Dana Farber Cancer Institute, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Harvard Stem Cell Institute, Boston, Massachusetts 02115, USA. ⁴Department of Chemistry and Chemical Biology and Department of Physics, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA. ⁵Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ⁶Center for Individualized Medicine, Department of Molecular Pharmacology & Experimental Therapeutics, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA. ⁷Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02140, USA.

*These authors contributed equally to this work.

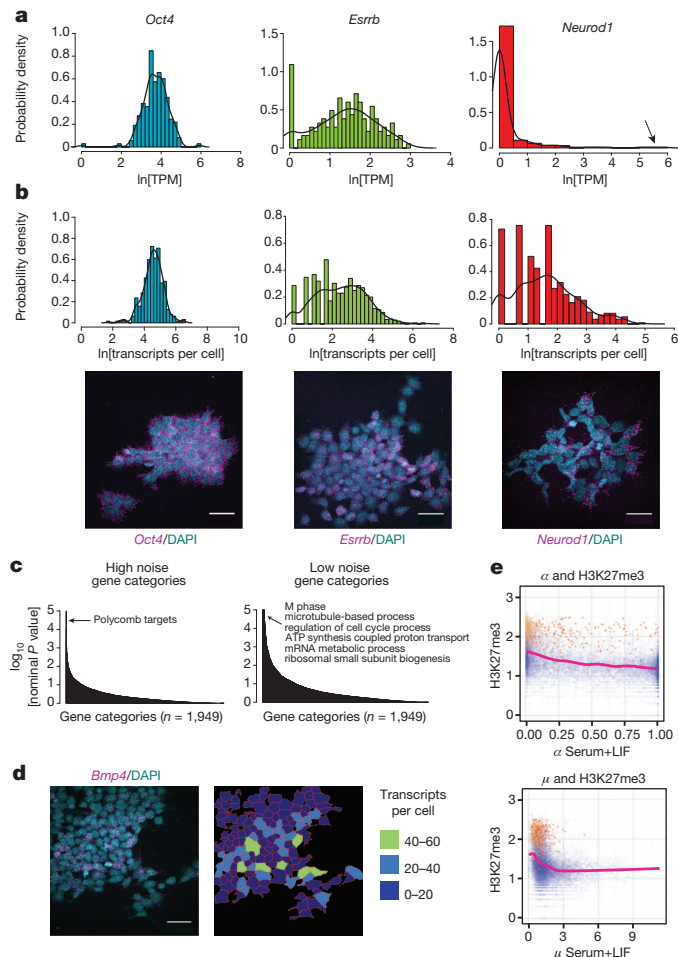


Figure 1 | Gene expression variability landscape of PSCs. **a**, Histograms of transcript distributions from single-cell RNA-seq of v6.5 mESCs cultured in serum + LIF. Arrow indicates high *Neurod1*-expressing cells. TPM, transcripts per million. **b**, Histograms and representative images of transcript distributions for *Oct4*, *Esrrb* and *Neurod1* from single-molecule FISH. Scale bars, 40 μ m. **c**, Gene categories showing high or low noise. **d**, Sporadic expression of the Polycomb target gene *Bmp4* within an mESC colony as measured by smFISH. **e**, Relationship between population H3K27me3 levels, fraction of cells a gene is detected in (α , top) and average expression level when detected (μ , bottom). Detected (red) and undetected (orange) Polycomb target genes are highlighted. Overall trend lines are shown. All relevant statistical information can be found in the 'Statistics' section of Methods. Scale bar, 40 μ m.

Certain Polycomb target genes were expressed in some cells at levels comparable to pluripotency regulators, up to 60 transcripts per cell, despite the presence of the repressive H3K27me3 (trimethylation of lysine 27 on histone H3) chromatin mark associated with Polycomb activity (Fig. 1d, Extended Data Fig. 4 and Supplementary Table 5). As expected, Polycomb target genes showed higher average levels of H3K27me3, were detected in a smaller fraction of cells, and were expressed at lower levels than non-Polycomb target genes (Fig. 1e). However, within this set of genes, those with detectable expression showed lower average levels of H3K27me3 than those that were not detected (Student's *t*-test $P = 4.82 \times 10^{-5}$), suggesting that dynamic fluctuations in chromatin state are associated with sporadic expression of certain Polycomb targets in PSCs. This subset of Polycomb target genes may represent regulators governing initial steps in lineage commitment, and may therefore be subject to particularly dynamic regulation in PSCs.

To determine the stability of expression states over rounds of cell division, we seeded individual mESCs onto culture plates, allowed them to form colonies over 3–4 days, and quantified the level of inter- and intra-colony variability for selected genes using smFISH (Fig. 2a). Individual

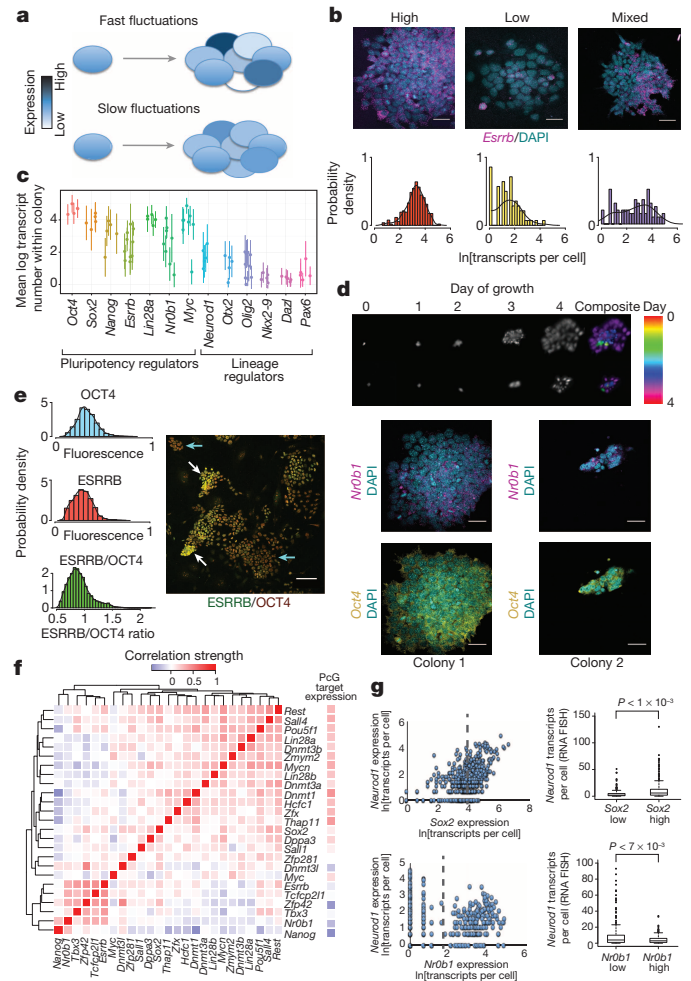


Figure 2 | Expression states of variable genes are coupled together and persist over multiple cell divisions. **a**, Slowly fluctuating genes show a high degree of intercolony variability. **b**, Expression of the pluripotency regulator *Esrrb* within individual colonies. **c**, Intra- and inter-colony variability in expression for selected pluripotency and lineage regulators. Average transcript number and standard deviation within colonies are indicated. **d**, Time-lapse imaging of colony formation from single cells, and *Nr0b1* and *Oct4* expression within these colonies. **e**, Relative OCT4 and ESRRB protein levels within mESCs cultured in serum + LIF. Groups of high and low ESRRB cells are indicated. **f**, Correlation of pluripotency regulator and Polycomb target gene expression between individual cells. PcG, Polycomb-group protein. **g**, Dependence of *Neurod1* expression on the level of *Sox2* and *Nr0b1* within individual cells. Dashed lines indicate high and low expression states, and *P* values for differences between states were calculated using the Kolmogorov–Smirnov test. Scale bars: **b**, **d**, 40 μ m; **e**, 0.01 cm.

colonies showed distinct expression states for *Esrrb*, with some locked into a high level, some locked into a low level, and others that displayed mixed expression suggesting a sudden switching between high and low states during the process of colony formation (Fig. 2b). Other pluripotency regulators, including *Nanog*, *Nr0b1* (also called *Dax1*) and *Myc*, as well as lineage regulators and Polycomb targets *Neurod1*, *Otx2*, *Olig2* and *Pax6*, also exhibited high inter-colony variability suggestive of slow fluctuations in expression states (Fig. 2c). As clusters of neighbouring cells or entire colonies tended to be in similar expression states (Extended Data Fig. 5), we estimate that transitions between transcriptional states for these variable regulators occur relatively infrequently with respect to the ES cell cycle, happening on the order of one to a few days, in line with measurements of *Nanog* transcriptional fluctuations^{6,9,28}. To confirm that 'variable expression' colonies were clonally derived, we performed time-lapse imaging to monitor colony formation over 4 days. Individual colonies formed from single cells showed substantial differences

in growth rate and bimodal *Nr0b1* expression (Fig. 2d), validating our approach and highlighting the pronounced variability and persistence of growth rate, morphology and expression state of mESCs grown in serum plus LIF (serum + LIF mESCs). This expression state persistence extended to the protein level (Fig. 2e and Extended Data Fig. 5), indicating that slow fluctuations in expression of certain pluripotency regulators might underlie distinct phenotypic responses of individual mESCs to external stimuli²⁹.

Clustering of pluripotency regulators revealed that they partitioned into several co-expressed modules, with some modules positively correlated with Polycomb target expression and others negatively correlated (Fig. 2f and Extended Data Fig. 5). *Nanog*, *Nr0b1* and *Zfp42* (also called *Rex1*) were among those showing the strongest negative correlation with Polycomb target gene expression. To test these associations, we examined the dependence of selected Polycomb target genes on individual pluripotency regulators by two-colour FISH (Fig. 2g and Extended Data Fig. 5). Consistent with the RNA-seq data, *Neurod1* was more likely to be expressed at high levels in cells expressing high levels of *Sox2* (which also functions to specify the neural lineage) and low levels of *Nr0b1*, while *Bmp4* expression was positively associated with *Esrrb*, and *Otx2* expression was negatively associated with *Nanog*. Notably, the 8% of cells separated from the main population by principal component analysis tended to be in low expression states for *Nr0b1*, *Nanog* and *Zfp42*, and showed higher Polycomb target gene expression than did the majority of cells (Extended Data Figs. 2 and 5), suggesting that the RNA-seq-detected correlations reflect biologically meaningful states in which pluripotency factor expression influences the probability of lineage regulator expression.

Regulatory pathways govern PSC heterogeneity and state

For a better understanding of the factors governing PSC heterogeneity, we examined mouse PSCs cultured under different growth conditions¹³, treated with chemical inhibitors of epigenetic regulators, genetically modified to lack particular regulators^{30–34}, or in different states of pluripotency³⁵ (Extended Data Table 1, Supplementary Table 6 and Extended Data Fig. 6). While unimodally expressed genes such as *Oct4* and *Rest* remained relatively invariant across a range of perturbations, the distributions of bimodally expressed genes such as *Myc* and *Tcfcp2l1* shifted markedly in response to particular perturbations (Fig. 3a). Culturing mouse PSCs with inhibitors of Erk and GSK3 signalling ('2i' or ground state conditions) has previously been shown to promote self-renewal and block differentiation¹³. Notably, both culturing mESCs in 2i + LIF conditions¹³ and impairing microRNA (miRNA) production^{30,34} resulted in more uniform gene expression across 15 pluripotency regulators examined by single-cell quantitative polymerase chain reaction (qPCR), whereas knocking out PRC2 function through loss of the Polycomb-group (PcG) protein EED (ref. 31) resulted in greater population heterogeneity across all genes measured (Fig. 3b).

We next performed principal component analysis on the single-cell qPCR data and applied an automated classification algorithm to assign cells to discrete states (Fig. 3c–e and Extended Data Fig. 6). Surprisingly, a large fraction of both *Dgcr8*^{−/−} and *Dicer* knockout mESCs, which lack mature miRNAs due to loss-of-function of separate miRNA processing factors^{30,34}, were assigned to the ground state, suggesting that blocking external signalling pathways and removing miRNAs from PSCs results in common modes of self-renewal (Fig. 3e). PSCs in most other conditions were predominantly classified as a distinct state, which

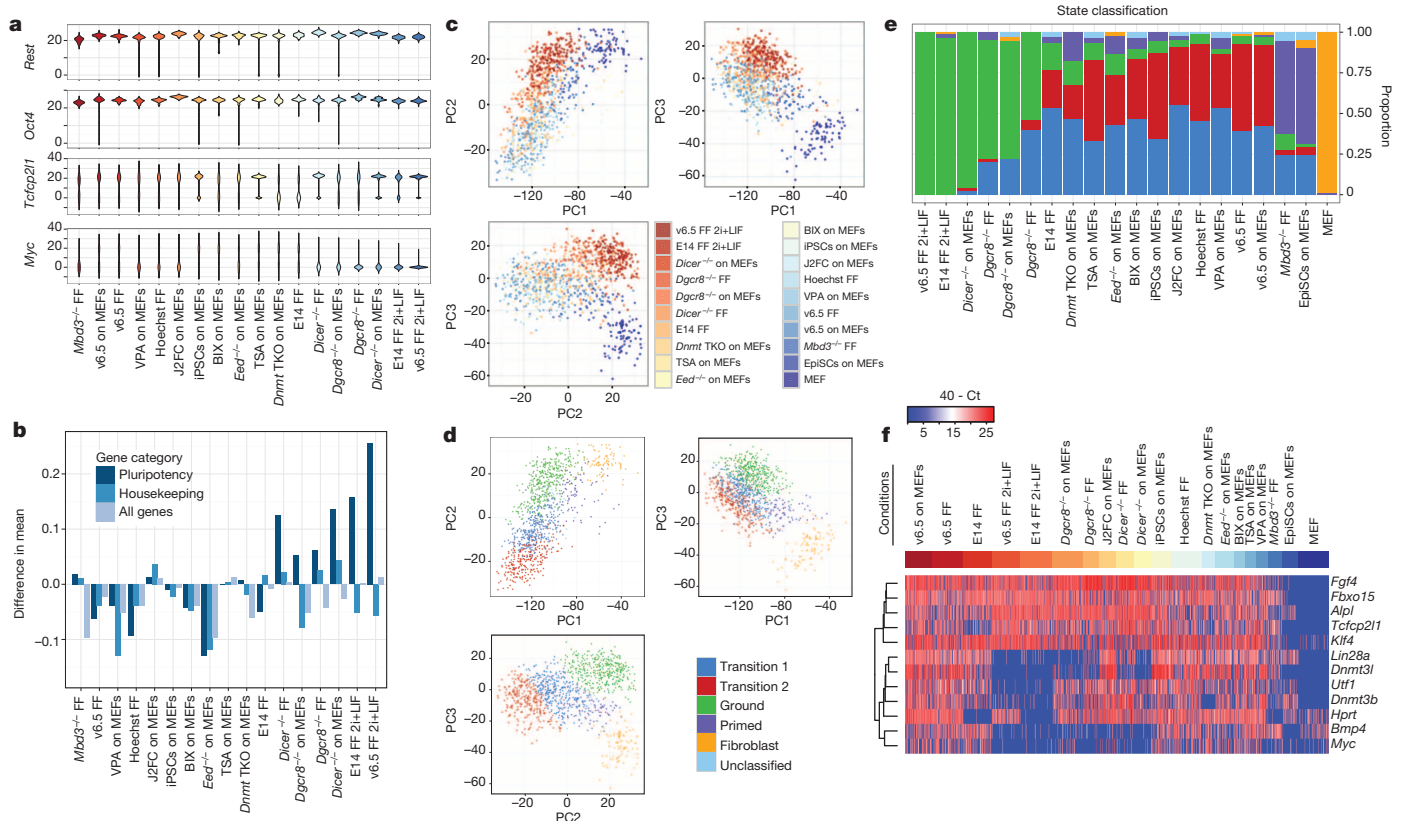


Figure 3 | Effect of perturbations on gene expression variability and cell state. **a**, Population distributions for the unimodally expressed genes *Oct4* and *Rest*, and the bimodally expressed genes *Tcfcp2l1* and *Myc*. iPSCs, induced pluripotent stem cells; MEFs, mouse embryonic fibroblasts; TKO, triple knockout. See Extended Data Table 1 for a full list of conditions and abbreviations. **b**, Comparison of population heterogeneity as measured by

difference in mean intra-condition distance between serum + LIF and listed condition. **c**, Principal component analysis (PCA) of single-cell qPCR data. **d**, PCA coloured by the most likely state classification. **e**, Cell state classification of PSCs exposed to different perturbations and conditions. **f**, Expression heat map of genes contributing the most to the top three principal components, excluding housekeeping and fibroblast genes.

we term the 'transition state', as it is primarily associated with cells cultured under serum + LIF conditions and a higher probability of spontaneous differentiation. Epiblast stem cells (EpiSCs) were assigned to a separate state we term 'primed', in keeping with the view of these cells as being more developmentally advanced³⁵. Genes contributing the most to the principal components that distinguished *Dgcr8*^{-/-} and *Dicer* knockout mESCs in serum + LIF and wild-type mESCs in 2i + LIF from mESCs in other conditions included *Myc*, *Lin28a*, *Bmp4*, *Dnmt3b* and *Dnmt3l*, all of which showed sharply reduced expression in ground state cells (Fig. 3f). Our perturbation analysis therefore implicates miRNAs as key mediators of the transition and primed states, with their absence mimicking the inhibition of the Erk and GSK3 signalling pathways observed in 2i culture.

2i and miRNA deficiency promote the PSC ground state

To investigate further the apparent similarity between wild-type mESCs cultured in 2i + LIF conditions (2i + LIF mESCs) and *Dgcr8*^{-/-} mESCs cultured in serum + LIF (*Dgcr8*^{-/-} mESCs), we performed single-cell RNA-seq on mESCs from each condition (Supplementary Tables 2, 7, 8, 9). Gene expression changes between serum + LIF and 2i + LIF mESCs were highly correlated with changes between serum + LIF and *Dgcr8*^{-/-} mESCs (Fig. 4a and Extended Data Fig. 7). As a population, 2i + LIF mESCs showed reduced heterogeneity as compared to serum + LIF mESCs, in keeping with the notion that inhibiting Erk and GSK3 signalling drives mouse PSCs into a low-noise ground state^{20,28,37} (Fig. 4b and Extended Data Fig. 7). By contrast, *Dgcr8*^{-/-} mESCs displayed increased population heterogeneity, consistent with a role for miRNAs in buffering gene expression noise³⁸. When compared across conditions, however, individual *Dgcr8*^{-/-} mESCs were more similar to 2i + LIF than serum + LIF mESCs, supporting a model in which removal of miRNAs drives a portion of *Dgcr8*^{-/-} mESCs towards the ground state, and results in a similar phenotype to that produced by Erk and GSK3 inhibition.

While some pluripotency regulators such as *Esrrb* and *Tbx3* showed increased expression in 2i + LIF and *Dgcr8*^{-/-} mESCs, other factors traditionally associated with pluripotency (for example, *Myc* and *Lin28a*) showed a sharp reduction in expression under both conditions (Fig. 4c). DNA methyltransferases *Dnmt3a*, *Dnmt3b* and *Dnmt3l* showed reduced

expression in both 2i + LIF and *Dgcr8*^{-/-} mESCs, suggesting that miRNA activity may be linked to the reduction in DNA methylation observed in the naive pluripotent state^{28,39}. Regulators displaying bimodal expression patterns in serum + LIF mESCs, including the pluripotency factor *Esrrb*, showed altered distributions in both 2i + LIF and *Dgcr8*^{-/-} mESCs. Expression was confirmed by smFISH and quantitative immunofluorescence, and was recapitulated in an independent mESC line (Fig. 4d, e and Extended Data Fig. 7). Thus, the pluripotency regulatory network adopts distinct configurations in the ground and transition states.

mESCs cultured in 2i + LIF exhibit a unique chromatin state characterized by lower levels of H3K27me3 at promoters and increased amounts of transcriptional pausing⁴⁰. Both culture in 2i + LIF and removal of mature miRNAs resulted in a reduction of H3K27me3 at promoters (Fig. 4f, g, Extended Data Fig. 8 and Supplementary Table 5). Therefore, *Dgcr8*^{-/-} mESCs manifest features of ground-state self-renewal including altered gene expression, a reconfigured pluripotency regulatory network, and a common chromatin state.

miRNA balance and cell state

We profiled miRNA expression in two mESC lines cultured under both serum + LIF and 2i + LIF conditions, and mouse embryonic fibroblasts (MEFs) cultured under standard conditions (Supplementary Table 10). Notably, while levels of ES-cell-specific cell-cycle-regulating (ESCC) miRNAs³⁶, which are known to be expressed at high levels in PSCs, remained elevated in both serum + LIF and 2i + LIF growth conditions, let-7 family members and miR-152, which act as tumour suppressors and are typically associated with differentiated cells^{41–43}, were expressed at higher levels in 2i + LIF as compared to serum + LIF culture (Fig. 5a and Extended Data Fig. 9).

Prevailing models which posit that the ESCC and let-7 miRNA families are opposing classes of miRNAs that act in self-reinforcing loops to stabilize self-renewing and differentiated states⁴⁴ run contrary to our observed co-expression of ESCC and let-7 miRNAs in 2i + LIF mESCs. Analysing our own and published data⁴⁰, we found that targets of let-7 showed significantly lower expression as a group in 2i + LIF as compared to serum + LIF mESCs ($P = 1.4 \times 10^{-5}$, Mann–Whitney *U*-test), as did miR-152 predicted targets ($P = 0.005$), suggesting that these miRNAs

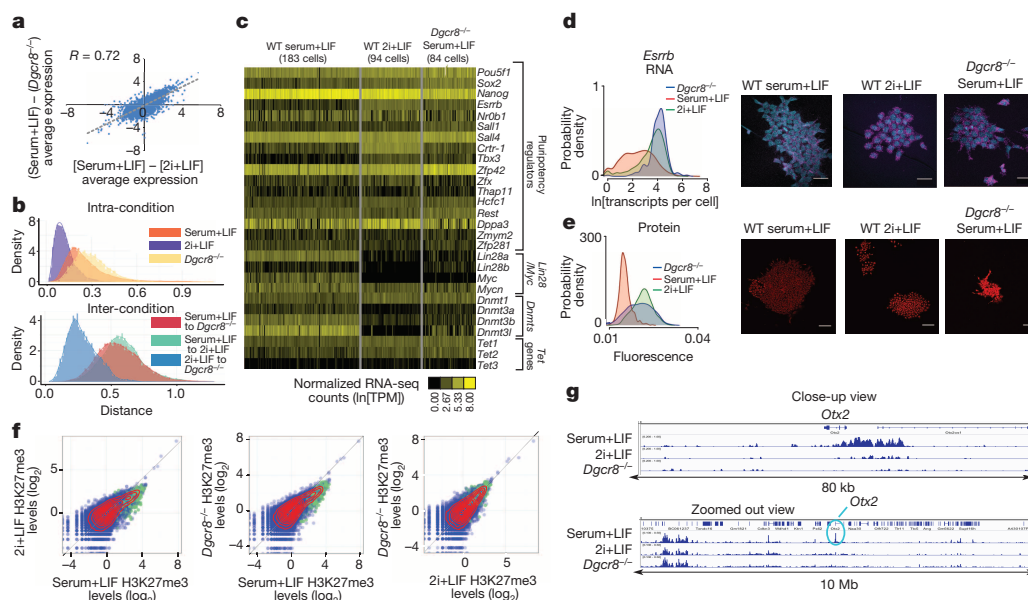


Figure 4 | *Dgcr8*^{-/-} mESCs show evidence of ground-state self-renewal.

a, Correlation between single-cell RNA-seq gene expression changes in different conditions. **b**, Distances between individual cells for pluripotency regulators shown in panel **c**. **c**, Heat map of single-cell RNA-seq data for selected pluripotency regulators. **d**, **e**, Single-molecule FISH (**d**) and quantitative immunofluorescence (**e**) showing a shift towards the high

expression state of *Esrrb* in 2i + LIF and *Dgcr8*^{-/-} mESCs. Scale bars: 40 μm (top row); 100 μm (bottom row). **f**, Promoter H3K27me3 levels in the three conditions examined by single-cell RNA-seq. Polycomb target genes are shown in green. **g**, H3K27me3 ChIP-seq tracks from the three conditions profiled showing the selective loss of H3K27me3 at the *Otx2* promoter in *Dgcr8*^{-/-} and 2i + LIF mESCs.

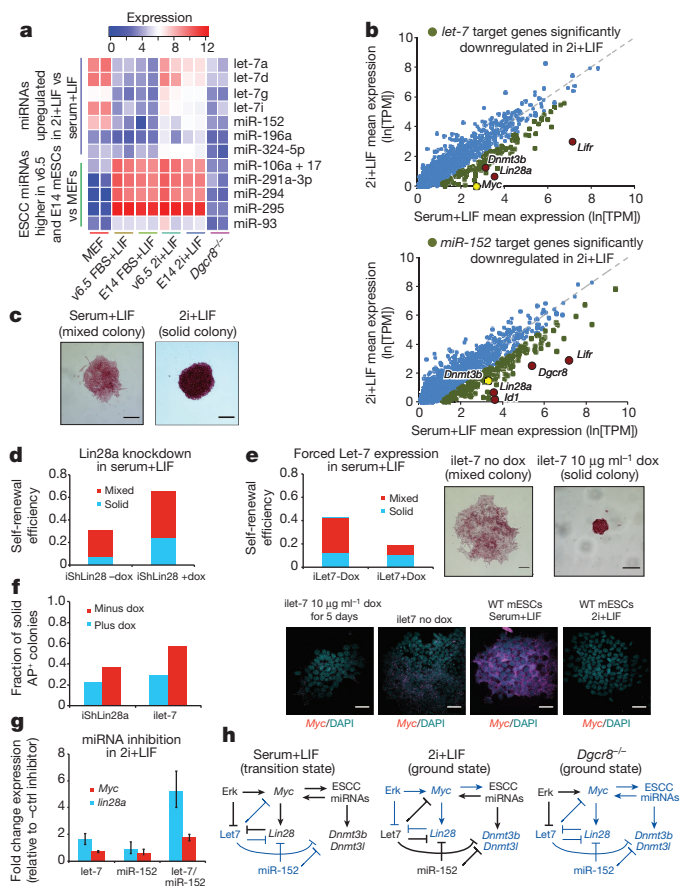


Figure 5 | miRNA balance controls transitions between ground and transition states. **a**, NanoString profiling of miRNAs expressed in mESCs or MEFs. **b**, Expression changes of predicted let-7 or miR-152 target genes between conditions. **c**, Representative colonies showing solid or mixed alkaline phosphatase staining. **d**, **e**, Self-renewal efficiency of mESCs bearing a doxycycline-inducible *Lin28a* shRNA construct (**d**) or doxycycline-inducible let-7 (**e**) in the presence or absence of doxycycline. Mean values from two replicate experiments are shown. *Myc* expression levels for representative iLet-7 colonies as measured by smFISH are shown in **e**. Scale bars: 100 µm (top row); 40 µm (bottom row). **f**, Fraction of uniformly alkaline-phosphatase-positive (AP⁺) colonies in experiments shown in **d** and **e**. **g**, Expression changes in *Myc* and *Lin28a* induced by transfection of miRNA inhibitors into wild-type mESCs cultured in 2i+LIF. Error bars indicate standard deviations between triplicate transfection experiments. **h**, Model for interplay between Erk signalling, miRNAs and *Myc*/*Lin28a*/let-7 axis in ground and transition states.

function to repress a set of target genes under ground-state conditions (Supplementary Table 9). Genes targeted by either or both let-7 and miR-152 that were downregulated in 2i+LIF as compared to serum+LIF included *Myc*, *Lin28a*, *Lin28b* and *Dnmt3b* (Fig. 5b). Furthermore, predicted let-7 and miR-152 targets were more highly correlated in expression at a single-cell level in 2i+LIF as compared to non-target genes (Supplementary Information), suggesting that these genes and miRNAs comprise a distinct regulatory module.

Enforced expression of ESCC miRNAs can correct cell cycle defects of *Dgcr8*^{-/-} mESCs, promote rapid proliferation of these cells, suppress inhibitors of the G1 to S transition, and upregulate genes including *Myc* and *Lin28* through presumed indirect mechanisms (refs 36, 44). Transfection of the prototypical ESCC miRNA miR-294 into *Dgcr8*^{-/-} mESCs has been shown to upregulate *Myc*, *Lin28*, *Dnmt3b* and *Dnmt3l* (ref. 44), which are let-7/miR-152 target genes suppressed in the ground state (Extended Data Fig. 9). This opposing effect of ESCC and let-7 miRNAs on a common set of target genes suggests a role for let-7 during differentiation, where it acts to repress pluripotency factors sustained by ESCC miRNAs in stem cells⁴⁴. However, despite the continued high

expression of ESCC miRNAs in 2i, their direct targets showed elevated expression as a class in 2i as compared to serum ($P = 6.412 \times 10^{-9}$, Mann–Whitney *U*-test, Extended Data Fig. 9), whereas let-7 and miR-152 direct targets were downregulated in 2i. These findings suggest that enforced expression of ESCC miRNAs elevates *Dgcr8*^{-/-} mESCs out of ground-state self-renewal through indirect modulation of the *Myc*/*Lin28*/let-7 axis, and that in 2i let-7 and miR-152 oppose this effect of ESCC miRNAs through direct repression of a set of target genes including *Myc*, *Lin28*, *Dnmt3b* and *Dnmt3l*.

Both 2i+LIF and *Dgcr8*^{-/-} mESCs give rise predominantly to compact colonies uniformly positive for alkaline phosphatase staining, as compared to serum+LIF mESCs which have a greater tendency to form larger colonies that are mixed for alkaline phosphatase staining (Fig. 5c). Compact, uniformly alkaline-phosphatase-positive colonies formed in serum+LIF culture displayed lower levels of *Myc* expression than did alkaline phosphatase mixed and negative colonies (Extended Data Fig. 9). We independently knocked down *Lin28a* expression in serum+LIF mESCs, and forced let-7 expression in doxycycline-inducible iLet-7 mESCs grown in serum+LIF⁴⁵. *Lin28a* blocks processing of let-7 precursor miRNAs into their mature form, and knockdown of *Lin28a* in mESCs grown in serum+LIF results in increased let-7 levels, increased colony-forming efficiency, and a higher proportion of smaller, uniformly alkaline-phosphatase-positive colonies (Fig. 5d and Extended Data Fig. 9). Sustained expression of let-7 in serum+LIF culture using the inducible system resulted in a higher proportion of smaller, uniformly alkaline-phosphatase-positive colonies that showed lower levels of *Myc* expression (Fig. 5e, f and Extended Data Fig. 9), suggesting that let-7 is capable of activating a positive feedback circuit that stabilizes ground state self-renewal. Acute inhibition of let-7 family members and miR-152 in wild-type mESCs cultured in 2i+LIF resulted in upregulation of *Myc* and *Lin28a* (Fig. 5g), further supporting the notion that these two miRNAs act in concert to maintain the ground state.

We tested the effect of reintroducing miRNAs into *Dgcr8*^{-/-} mESCs on their self-renewal to see if ESCC-mediated elevation into the transition state could be counterbalanced by let-7. Transfection of stable mimics of let-7 resulted in sharply reduced self-renewal efficiency compared to controls, and transfection of mimics of the ESCC miRNA miR-294 also resulted in reduced self-renewal (Extended Data Fig. 9). Co-transfection of both miRNA family members together, however, resulted in higher self-renewal efficiency than did introduction of either family member alone. These results support a model in which expression of ESCC miRNAs alone drives PSCs into a transition state with a relatively high probability of spontaneous differentiation through indirect activation of *Myc* and *Lin28a*, while expression of let-7 miRNAs alone can drive differentiation. However, expression of a balance of ESCC miRNAs along with the opposing families let-7 and miR-148/152 as in 2i conditions, or neither as in the case of *Dgcr8*^{-/-} mESCs, results in ground-state self-renewal through either direct repression of *Myc* and *Lin28a* by let-7 and miR-148/152, or removal of the activating force of ESCC miRNAs (Fig. 5h).

Conclusion

The diverse range of conditions under which pluripotency can be induced or maintained has been accompanied by reports of molecular and functional variation. Here we analysed the dynamic transcriptional landscape of pluripotent stem cells subject to a number of chemical and genetic perturbations. Applying single-cell analytics, we gleaned a number of essential insights. We found that different classes of genes manifest high or low expression variability in PSCs, with housekeeping and metabolic gene sets showing consistent expression across individual cells, while genes involved in signalling pathways and development were considerably more variable. Moreover, expression states of variable regulatory factors were coupled together, implying the presence of a regulated biological network. Analysis of chemical and genetic perturbations led to the discovery that depletion of miRNAs mimicked the transcriptional ground state of pluripotency routinely induced by culture in 2i+LIF,

conditions that block the dominant ERK and GSK3 signalling pathways that converge on the *Myc/Lin28/let-7* axis. Our data shed light on the transcriptional dynamics of the pluripotent state at the single-cell level, and demonstrate how regulation of gene expression variation relates directly to the transition between pluripotency and differentiation. Transcriptional heterogeneity is increasingly being recognized as a key component of many biological processes^{46–48}. It will be of interest to map stable and flexible regulatory nodes in networks governing other progenitor and differentiated cell types to discern common principles underlying network architecture and gene expression variability.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 February; accepted 7 October 2014.

- Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
- Loh, Y.-H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38**, 431–440 (2006).
- Loh, Y. H. *et al.* Genomic approaches to deconstruct pluripotency. *Annu. Rev. Genomics Hum. Genet.* **12**, 165–185 (2011).
- MacArthur, B. D., Ma'ayan, A. & Lemischka, I. R. Systems biology of stem cell fate and cellular reprogramming. *Nature Rev. Mol. Cell Biol.* **10**, 672–681 (2009).
- Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011).
- Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234 (2007).
- Hayashi, K., Lopes, S. M. C. S., Tang, F. & Surani, M. A. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* **3**, 391–401 (2008).
- Hong, S.-H. *et al.* Cell fate potential of human pluripotent stem cells is encoded by histone modifications. *Cell Stem Cell* **9**, 24–36 (2011).
- Kalmar, T. *et al.* Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol.* **7**, e1000149 (2009).
- Karwacki-Neisius, V. *et al.* Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. *Cell Stem Cell* **12**, 531–545 (2013).
- MacArthur, B. D. *et al.* Nanog-dependent feedback loops regulate murine embryonic stem cell heterogeneity. *Nature Cell Biol.* **14**, 1139–1147 (2012).
- Reynolds, N. *et al.* NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell* **10**, 583–594 (2012).
- Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
- Arias, A. M. & Brickman, J. M. Gene expression heterogeneities in embryonic stem cell populations: Origin and function. *Curr. Opin. Cell Biol.* **23**, 1–7 (2011).
- Cherry, A. & Daley, G. Q. Another horse in the meta-stable state of pluripotency. *Cell Stem Cell* **7**, 641–642 (2010).
- Graf, T. & Stadtfeld, M. Heterogeneity of embryonic and adult stem cells. *Cell Stem Cell* **3**, 480–483 (2008).
- Halley, J. D. *et al.* Self-organizing circuitry and emergent computation in mouse embryonic stem cells. *Stem Cell Res.* **8**, 324–333 (2012).
- Loh, K. M. & Lim, B. A precarious balance: pluripotency factors as lineage specifiers. *Cell Stem Cell* **8**, 363–369 (2011).
- MacArthur, B. D. & Lemischka, I. R. Statistical mechanics of pluripotency. *Cell* **154**, 484–489 (2013).
- Silva, J. & Smith, A. Capturing pluripotency. *Cell* **132**, 532–536 (2008).
- Thomson, M. *et al.* Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* **145**, 875–889 (2011).
- Cahan, P. & Daley, G. Q. Origins and implications of pluripotent stem cell variability and heterogeneity. *Nature Rev. Mol. Cell Biol.* **14**, 357–368 (2013).
- Huang, S. Non-genetic heterogeneity of cells in development: more than just noise. *Development* **136**, 3853–3862 (2009).
- Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240 (2013).
- Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363–369 (2014).
- Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **4**, e1000242 (2008).
- Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
- Singer, Z. S. *et al.* Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell* **55**, 319–331 (2014).
- Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444**, 643–646 (2006).
- Calabrese, J. M., Seila, A. C., Yeo, G. W. & Sharp, P. A. RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **104**, 18097–18102 (2007).
- Chamberlain, S. J., Yee, D. & Magnuson, T. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem Cells* **26**, 1496–1505 (2008).
- Kaji, K. *et al.* The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nature Cell Biol.* **8**, 285–292 (2006).
- Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–814 (2006).
- Wang, Y., Medvid, R., Melton, C., Jaenisch, R. & Blueloch, R. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nature Genet.* **39**, 380–385 (2007).
- Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–199 (2007).
- Wang, Y. *et al.* Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nature Genet.* **40**, 1478–1483 (2008).
- Grün, D., Kester, L. & van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nature Methods* **11**, 637–640 (2014).
- Ebert, M. S. & Sharp, P. A. Roles for microRNAs in conferring robustness to biological processes. *Cell* **149**, 515–524 (2012).
- Leitch, H. G. *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nature Struct. Mol. Biol.* **20**, 311–316 (2013).
- Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).
- Boyerinas, B., Park, S. M., Hau, A., Murmann, A. E. & Peter, M. E. The role of let-7 in cell differentiation and cancer. *Endocr. Relat. Cancer* **17**, F19–F36 (2010).
- Tsuruta, T. *et al.* miR-152 is a tumor suppressor microRNA that is silenced by DNA hypermethylation in endometrial cancer. *Cancer Res.* **71**, 6450–6462 (2011).
- Reinhart, B. J. *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
- Melton, C., Judson, R. L. & Blueloch, R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* **463**, 621–626 (2010).
- Zhu, H. *et al.* The Lin28/let-7 axis regulates glucose metabolism. *Cell* **147**, 81–94 (2011).
- Balázs, G., van Oudenaarden, A. & Collins, J. J. Cellular decision making and biological noise: from microbes to mammals. *Cell* **144**, 910–925 (2011).
- Blake, W. J. *et al.* Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* **24**, 853–865 (2006).
- Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank members of the Collins and Daley laboratories for discussions. J.J.C. is supported by NIH grant R24DK092760 and the HHMI. G.Q.D. is supported by grants from the NIH (R01GM107536, R24DK092760, P50HG005550) and is an affiliate member of the Broad Institute and an investigator of the Manton Center for Orphan Disease Research and the Howard Hughes Medical Institute. A.R. is supported by the Broad Institute, the Klarman Cell Observatory at the Broad Institute, an NIH CECS (1P50HG006193-01), an NIH Pioneer Award (DP1OD003958-01) and the HHMI. R.M.K. is supported by the Wyss Institute. P.C. is supported by NIDDK (K01DK096013) and received support from NHLBI (T32HL066987 and T32HL007623) and the Manton Center for Orphan Disease Research. R.S. was supported by a NIH Postdoctoral Fellowship (1F32HD075541-01). H.L. is supported by the Mayo Clinic Center for Individualized Medicine. Sequencing was performed by the Broad Institute Genomics Platform. Flow cytometry was performed in the Hematologic Neoplasia Flow Cytometry Facility at the Dana-Farber Cancer Institute and the BCH IDDRC Stem Cell Core Facility at Boston Children's Hospital supported by NIH-P30-HD18655. Single-cell qPCR experiments were performed at the BCH IDDRC Molecular Genetics Core Facility at Children's Hospital Boston supported by NIH-P30-HD18655. Fluidigm C₁ experiments were performed at the Broad Institute and the Biopolymers Facility at Harvard Medical School.

Author Contributions R.M.K. designed and performed the experiments, analysed the data and wrote the paper. P.C. analysed the data, developed analytical tools and wrote the paper. A.K.S., D.G. and J.J.T. performed ChIP-seq and single-cell RNA-seq experiments. R.S. helped to analyse the ChIP-seq and single-cell RNA-seq data. A.D. performed experiments and helped to analyse the data. H.L. helped to analyse the data. J.Z. generated the iShLin28 mESCs. K.P. generated the iPSCs. T.C.F. assisted with imaging and wrote image analysis algorithms. A.R. oversaw single-cell RNA-seq experiments and helped to write the paper. J.J.C. and G.Q.D. oversaw the project and helped to design the study and write the paper.

Author Information Data are deposited in GEO under accession number GSE60749. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.Q.D. (george.daley@childrens.harvard.edu) or J.J.C. (jjcollins@bu.edu).

METHODS

Cell culture. Mouse embryonic stem cells (mESCs) and induced pluripotent stem cells (iPSCs) were cultured on γ -irradiated mouse embryonic fibroblasts (MEFs) (GlobalStem, CF-1 MEFs) or in feeder-free conditions as detailed in Extended Data Table 1. Feeder-free cells were grown on Matrigel (hESC-qualified, BD)-coated Nunc cell-culture treated plates unless otherwise specified. Cells cultured in Serum + LIF were grown in Dulbecco's Modified Eagle Medium (DMEM) + Glutamax (Gibco) plus 15% fetal bovine serum (FBS, Stasis stem cell-qualified, Gemini), 1 \times penicillin/streptomycin (Gibco), 1 \times non-essential amino acids (NEAA) (Gibco), 1 \times β -mercaptoethanol (β -Me) (EmbryoMax ES cell-qualified, Millipore), and 1,000 U ml⁻¹ leukaemia inhibitory factor (LIF) (ESGRO, Millipore). Cells cultured in 2i + LIF were maintained in N2/B27 media, consisting of 50% DMEM/F12 (1:1) and 50% Neurobasal plus 0.5 \times N-2 supplement, 0.5 \times B-27 supplement, 1 \times penicillin/streptomycin, 1 \times Glutamax, and 1 \times NEAA (all from Gibco), and 1 \times β -Me (Millipore) supplemented with 1 μ M MEK inhibitor PD0325901 (StemGent), 3 μ M GSK3 inhibitor CHIR99021 (StemGent), and 1,000 U ml⁻¹ LIF (Millipore). For the 1i experiments shown in Extended Data Fig. 9h, either the MEK inhibitor PD0325901 or the GSK3 inhibitor CHIR99021 were omitted from the media, respectively. Except as described, all mouse PSCs were passaged with trypsin (Gibco). Epiblast stem cells (EpiSCs) were grown on MEF feeder cells in DMEM-F12 (Gibco) supplemented with 20% knockout serum replacement (Life Technologies), 5 ng ml⁻¹ FGF2 (R&D Systems), 0.1 mM 2-mercaptoethanol (Sigma), 2 mM L-glutamine (Gibco), and 1 \times NEAA (Gibco), and passaged using collagenase type IV (Life Technologies).

MEFs used in the miRNA profiling experiments described in Fig. 5a were from GlobalStem (C57/BL6 MEFs), harvested at the second passage after thawing from frozen stocks, and cultured under standard conditions in DMEM + 10% FBS + 1 \times Pen/Strep/Glut.

For perturbation experiments described in Fig. 3, v6.5 mESCs were treated with 10 nM trichostatin A (Sigma), 1 mM valproic acid (Sigma), or 4 μ M BIX01294 (Stemgent) for 6 days before analysis, passaging twice in the interim.

Origin of cell lines used in this study. V6.5 mESCs were a gift from R. Jaenisch. E14TG2a mESCs were obtained from ATCC (CRL-1821). *Eed*^{-/-} mESCs were a gift from T. Magnuson. *Dnmt1*^{-/-} *Dnmt3a*^{-/-} *Dnmt3b*^{-/-} triple knockout mESCs were a gift from M. Okano. *Dgcr8*^{-/-} mESCs were obtained from Novus (NBA1-19349). *Mbd3*^{-/-} mESCs were a gift from B. Hendrich. mEpiSCs were a gift from P. Tesar. *Dicer* knockout mESCs and their matched J1 wild-type control mESCs were a gift from P. Sharp. iPSCs and iShLin28a cells were derived as described below.

Generation of iPSCs. Lentivirus production and generation of the iPSCs profiled by single-cell qPCR were performed as previously described⁴⁹. Briefly, iPSCs were generated from C57BL/6 MEFs (GlobalStem) using the doxycycline-inducible Tet-STEMCCA cassette and a constitutive reverse tetracycline transactivator (rtTA, gifts of G. Mostoslavsky) packaged into lentiviral vectors. In the presence of doxycycline the Tet-STEMCCA cassette expresses OCT4, KLF4, SOX2 and Myc from a single multicistronic transcript. The MEFs (P2) were plated at 25,000 cells ml⁻¹ into a gelatin-coated, six-well plate and allowed to become adherent overnight. Cells were then washed with PBS and transferred into 2 ml of retroviral infection medium⁵⁰ containing 25 μ l of concentrated Tet-STEMCCA and rtTA lentiviruses and 5 μ g ml⁻¹ polybrene, and incubated for 24 h. The virus-containing medium was then removed and replaced with 2 ml of fresh ES cell medium containing 1 μ g ml⁻¹ doxycycline, which was changed daily until the formation of colonies. Colonies were then picked and characterized for pluripotency markers.

Generation of inducible shLin28a ES cell line. The miR-30-based shRNA for Lin28a was designed as described⁵¹, and cloned into a cTGM targeting construct that is integrated at single copy into KH2 ES cells by Flp-driven recombinase-mediated cassette exchange. The KH2 cells contain a recipient locus downstream of *Col1a1* gene, and an rtTA expressed from the *Rosa26* promoter, allowing GFP expression and gene silencing upon addition of 1 μ g ml⁻¹ doxycycline.

Derivation of neural precursor cells from mESCs. V6.5 mESCs were differentiated to nestin-positive neural precursor cells by established methods^{52,53}. Briefly, feeder-free mESCs were allowed to form embryoid bodies for 4 days in suspension culture by plating on a Petri dish in serum + LIF mESC media. At day 4, embryoid bodies were plated on to a tissue culture surface and transferred to DMEM/F12 medium supplemented by insulin, transferrin, selenium and fibronectin (ITSFn) the following day. Cells were maintained in ITSFn for 8 days to select for nestin-positive cells, and were then plated on to laminin and polyornithine-coated plates and shifted to N2 media supplemented with bFGF/laminin to expand nestin-positive neural precursor cells. Media was changed every 2 days, and cells were passaged with trypsin. Expanded neural precursor cells were harvested and processed for single-cell RNA-seq as described below.

Cell capture and whole transcriptome amplification. Single-cell whole transcriptome amplification (WTA) was performed using the Fluidigm C₁ Single-Cell Auto Prep System (C₁ System) as per the manufacturer's recommendations (full details available at <http://www.fluidigm.com>). Briefly, mES or neural precursor cells

were taken directly from culture, trypsinized, washed in PBS, spun down and resuspended at a concentration of 3×10^5 cells ml⁻¹ of complete media, mixed 7:3 with C₁ suspension reagent (Fluidigm), and loaded onto C₁ Single-Cell Auto Prep chips (C₁ chips; Fluidigm). After loading, each of the cell isolation chambers on the C₁ chip was optically inspected for the presence of a cell. Subsequently, these cells were lysed and SMART-Seq⁵⁴ whole transcriptome amplified (WTA) products were prepared with the C₁ System using the SMARTer Ultra Low RNA kit for Illumina Sequencing (Clontech) and protocols provided by Fluidigm.

Library preparation and RNA-seq. WTA products were harvested from the C₁ chip, diluted to a concentration of 0.15 ng μ l⁻¹, and cDNA libraries were prepared using Nextera XT DNA Sample preparation reagents (Illumina) as per the manufacturer's recommendations, with minor modifications. Specifically, reactions were run at one-quarter the recommended volume, the tagmentation step was extended to 10 min, and the extension time during the PCR step was increased from 30 s to 60 s. After the PCR step, all 96 samples were pooled without library normalization, cleaned twice with 0.9 \times AMPure XP SPRI beads (Beckman Coulter), and eluted in buffer TE. The pooled libraries were quantified using a Quant-IT DNA High-Sensitivity Assay kit (Invitrogen) and examined using a high-sensitivity DNA chip (Agilent). Finally, samples were sequenced deeply using either a HiSeq 2000 or a HiSeq 2500 sequencer (25-bp paired-end reads). Single-cell samples were sequenced to an average depth of more than ~ 1 million aligned reads per cell, with full values for number and fraction of aligned reads provided in the Supplementary Tables.

RNA-seq of population controls. Population controls were generated by extracting total RNA (RNeasy plus Micro RNA kit; Qiagen), according to the manufacturer's recommendations. Subsequently, 1 μ l of RNA in water was added to 2 μ l of C₁ lysis reaction mix, thermocycled using cycling conditions I (as below). Next, 4 μ l of the RT Reaction Mix were added and the mixture was thermocycled using cycling conditions II (as below). Finally, 1 μ l of the total RT reaction was added to 9 μ l of PCR mix and that mixture was thermocycled using cycling conditions III (as below). Products were quantified, diluted to 0.15 ng μ l⁻¹ and libraries were prepared, cleaned and tested as above.

For the cell lysis mix we used the following composition and cycling conditions. C₁ loading reagent (stock concentration 20 \times ; volume 0.60 μ l), SMARTer Kit RNase Inhibitor (40 \times ; 0.30 μ l), SMARTer Kit 3' SMART CDS Primer II A (12 μ M; 4.20 μ l), SMARTer kit dilution buffer (1 \times ; 6.90 μ l); cycling conditions I: (a) 72 $^{\circ}$ C, 3 min, (b) 4 $^{\circ}$ C, 10 min, (c) 25 $^{\circ}$ C, 1 min.

For the reverse transcription (RT) reaction mix we used the following composition and cycling conditions. C₁ loading reagent (stock concentration 20.0 \times ; volume 0.45 μ l), SMARTer kit 5X First-Strand Buffer (RNase-Free) (5.0 \times ; 4.20 μ l), SMARTer kit dithiothreitol (100 mM; 0.53 μ l), SMARTer kit dNTP Mix (dATP, dCTP, dGTP and dTTP, each at 10 mM) (10 mM; 2.10 μ l), SMARTer kit SMARTer II A oligonucleotide (12 μ M; 2.10 μ l), SMARTer kit RNase inhibitor (40 \times ; 0.53 μ l), SMARTer kit SMARTScribe reverse transcriptase (100.0 \times ; 2.10 μ l); cycling conditions II: (a) 42 $^{\circ}$ C, 90 min, (b) 70 $^{\circ}$ C, 10 min.

For the PCR mix we used the following composition and cycling conditions. PCR water (35.2 μ l), 10 \times Advantage 2 PCR buffer (stock concentration 10.0 \times ; volume 5.6 μ l), 50 \times dNTP mix (10 mM; 2.2 μ l), IS PCR primer (12 μ M; 2.2 μ l), 50 \times Advantage 2 Polymerase mix (50.0 \times ; 2.2 μ l), C₁ loading reagent (20.0 \times ; 2.5 μ l); cycling conditions III: (a) 95 $^{\circ}$ C, 1 min; (b) 5 cycles of (i) 95 $^{\circ}$ C, 20 s, (ii) 58 $^{\circ}$ C, 4 min, (iii) 68 $^{\circ}$ C, 6 min; (c) 9 cycles of (i) 95 $^{\circ}$ C, 20 s, (ii) 64 $^{\circ}$ C, 30 s, (iii) 68 $^{\circ}$ C, 6 min; (d) 7 cycles of (i) 95 $^{\circ}$ C, 30 s, (ii) 64 $^{\circ}$ C, 30 s, (iii) 68 $^{\circ}$ C, 7 min; (e) 72 $^{\circ}$ C, 10 min.

Calculation of expression levels and statistics. Mapping of reads and calculation of expression values was performed as described²⁴. Briefly, a Bowtie⁵⁵ index was created based on the UCSC knownGene⁵⁶ transcriptome, and paired-end reads were directly aligned to this index. RSEM⁵⁷ was run on these alignments to generate expression level estimates, and these estimates (τ) were multiplied by 1,000,000 to yield estimates of transcripts per million (TPM) for each gene. These TPM estimates were then transformed into log space by taking $\ln[\text{TPM} + 1]$, which we hereafter refer to as $\ln[\text{TPM}]$. Unless otherwise specified, all RNA-seq expression levels referred to in the manuscript represent \ln -transformed TPM values.

For the fraction of cells a gene was detected in, we considered as detected any gene showing an $\ln[\text{TPM}]$ value > 0 after filtering out outlier cells (see below). Average single-cell expression levels were calculated by first averaging TPM levels in linear space, and then log-transforming by taking $\ln[(\text{linear average TPM}) + 1]$. Gene expression levels in expressing cells, as shown in Fig. 1e, were calculated by averaging TPM levels in log space but only over the number of cells in the population that the gene was detected in. Standard deviations (σ) were calculated using gene expression values in log space, and coefficients of variation were calculated by dividing the population standard deviation by the average single-cell expression value as defined above.

We used Sailfish to estimate the abundances of non-coding RNAs in the single-cell RNA-seq libraries⁵⁸. Sailfish was executed using a transcript index file (generated as described below) and fastq files of the library reads to generate transcript level

abundances, which we summed on a per gene basis to generate gene level estimates. To generate a transcript index file, we used Ensembl BioMart (accessed on 4 June 2014), using the Ensembl Genes 75, GRCm38.2 protein_coding, miRNA, and lincRNA annotations, using a kmer size = 17 when compiling the Sailfish index. We used the same process to identify reliably detected genes using Sailfish-processed data as described in Supplementary Information section 'Single-cell RNA-seq reproducibility and sensitivity'. We judged genes detected in at least 21% of cells, or showing a maximum Sailfish expression level of $\ln[\text{TPM}] \geq 1.83$ and detected in at least 5% of cells, to be reliably detected.

Assessment of library quality and exclusion of cells from analysis. Library quality metrics, including genomic mapping rates, were calculated using PicardTools version 1.42 (<http://broadinstitute.github.io/picard/>). Barcoded RNA-seq libraries were inspected for alignment rates and number of genes detected, and criteria for removal of outlier cells were empirically determined from a combination of these factors. A large majority of cells grouped together with similar numbers of genes detected and alignment rates (Extended Data Fig. 1). Cells showing substantially fewer numbers of genes detected and/or lower alignment rates than these groups were filtered out, leaving 183 individual wild-type mESCs in serum + LIF, 94 individual wild-type mESCs in 2i + LIF, 84 *Dgcr8*^{-/-} mESCs in serum + LIF, and 55 ESC-derived neural precursor cells that passed this quality control step and were used in subsequent analyses.

Definition of Polycomb target genes for expression comparisons. Polycomb target genes in mESCs were defined as those called bound by the Polycomb repressive complex 2 (PRC2) subunits SUZ12 and EZH2, the Polycomb repressive complex 1 (PRC1) subunit Ring1B, and carrying the H3K27me3 mark in ref. 26. Those studies were performed in v6.5 mESCs grown under standard conditions in the presence of serum + LIF.

Identification of gene sets enriched for variable genes. We took a permutation approach to determine whether gene sets (Gene Ontology Biological Process annotations plus a list of Polycomb target genes identified in ref. 26) tended to have coefficients of variation (COVs) higher or lower than expected by chance. First, we computed the average COV of all genes in a gene set. Then, to determine whether it was higher or lower than expected by chance, we randomly sampled x genes where x is the number of genes in the gene set being assessed, and computed the average COV. We repeated this procedure 10,000 times, and computed the empirical P value as $1 + \frac{\text{number of random samples in which the mean COV} \geq \text{observed COV}}{1 + 10,000}$. Similarly, P values for gene sets having lower COV than expected by chance were computed as $1 + \frac{\text{number of random samples in which the mean COV} \leq \text{observed COV}}{1 + 10,000}$. Because COV depends on μ , and genes in some gene sets have high or low μ (for example, housekeeping genes and lineage regulators, respectively), we weighted the gene sampling such that the randomly sampled gene sets had equivalent μ distributions as the gene set being assessed. These P values were adjusted by applying Holm's method to correct for multiple tests. Only significant gene sets are listed in Fig. 1c even though this panel shows the nominal P values.

Correlation and distance analyses. The Pearson correlation coefficient was used to quantify gene expression correlation in Fig. 2f. All dendrograms were generated by unsupervised hierarchical clustering using $1 - \text{Pearson correlation coefficient}$ as the distance metric and average linkage as the linkage method. All intra- and inter-condition distances (Figs 3b and 4b and Extended Data Figs 4g and 7c) were computed as $1 - \text{Pearson correlation coefficient}$. For Fig. 3b, pluripotency genes examined were *Pou5f1* (*Oct4*), *Nanog*, *Fgf4*, *Rex1*, *Sall4*, *Dnmt1*, *Tcfp2l1*, *Zfx*, *Lin28a*, *Dnmt3l*, *Dnmt3b*, *Ronin*, *Stella*, *Sall1* and *Myc*, and housekeeping genes were *Actb*, *Gapdh*, *Rps6* and *Ppia*.

Identification of differentially expressed genes. To identify differentially expressed genes between mESCs cultured in serum + LIF, in 2i + LIF, or *Dgcr8*^{-/-} mESCs, we first quantile normalized $\ln[\text{TPM}]$ expression values across all three data sets. We performed pairwise comparisons of all three conditions, identifying differentially expressed genes using Student's t -test and adjusting P values with Holm's method to correct for multiple tests. We defined genes as differentially expressed as those with corrected P values < 0.01 and $|\log_2(\text{ratios})| > 1$. To identify genes differentially expressed between mESCs cultured in serum + LIF and ESC-derived neural precursor cells, we quantile normalized raw $\ln[\text{TPM}]$ expression values across these two data sets, and then identified differentially expressed genes as above.

RNA fluorescence in situ hybridization (FISH). For RNA FISH experiments, mESCs were seeded in the absence of feeder cells at a density of 200–500 cells per well of a Matrigel (BD)-coated 24-well glass bottom plate (MatTek, No. 1.5 glass, 13 mm diameter). This seeding density was empirically determined to yield sparse individual cells attaching to the matrigel-coated glass. Individual cells were allowed to form colonies over 3–4 days of culture, and then fixed with 4% formaldehyde for 30 min at room temperature and stained with RNA FISH probes (Panomics, QuantiGene ViewRNA ISH Cell Assays) according to manufacturer's recommendations, except that protease treatment was omitted. After staining with FISH probes, cells

were treated with $1 \times$ DAPI and 1:100 CellMask Green stain (Life Technologies) for 10 min at room temperature, washed, treated with ProLong Gold (Life Technologies), covered with a cover slip and allowed to cure overnight. Individual colonies were then identified manually and entire colonies or portions of them were scanned on a Leica SP5 X MP inverted confocal microscope outfitted with a tunable white light laser at $63 \times$ magnification (1.3 NA, glycerol immersion) in x , y and z using HyD detectors set to photon counting mode. z -intervals were spaced $0.3 \mu\text{m}$ apart, and colonies were scanned to their upper and lower limits resulting in on average 50–60 slices per colony. Five-channel imaging was performed to obtain brightfield, DAPI (ex 405 nm), CellMask Green (ex 488 nm), probe 1 (ex 564 nm), and probe 2 (ex 647 nm, if applicable) images.

To count numbers of transcripts and assign them to individual cells, we first generated two-dimensional projections of the three-dimensional stacks for the FISH probe channels, and then determined the number of transcripts per colony in Matlab using a modified version of a previously described analysis package⁵⁹. Briefly, numbers of spots were counted at a range of threshold values, spot count versus threshold plots were manually inspected to determine a threshold plateau range over which transcript number was relatively insensitive to the threshold value, and this was set as the threshold to use. These thresholds were then applied in CellProfiler⁶⁰, which was used to identify individual cells and assign transcripts to them. Nuclei were identified from two-dimensional projections of DAPI images in a semi-automated process, with manual editing as needed for nuclei that CellProfiler was not able to correctly identify and segment. Identified nuclei were then used as seeds to propagate outwards and identify cell boundaries on the basis of two-dimensional CellMask Green projections. Transcript spots were identified from images thresholded with the values obtained in Matlab, and then assigned to individual cells. The number of transcripts identified by this method agreed well with the number found by three-dimensional spot-finding in Matlab, indicating that the low background of confocal microscopy (particularly when using the HyD detectors in photon counting mode) and relatively flat morphology of the ESC colonies grown in the absence of feeders enables two-dimensional transcript quantification.

For all RNA FISH distributions shown, kernel density plots, which are estimates of the probability densities of the underlying distributions, were generated using the *density* function in R with default settings. Where histograms are shown, counts were binned into 20–60 bins and histograms were plotted using the *hist* function in R and overlaid with kernel densities.

Live cell tracking experiments. For the live cell tracking experiments shown in Fig. 2d, cells were plated and cultured as described above, except that cells were stained with $1 \mu\text{M}$ CellTrace calcein AM red-orange (Life Technologies), allowed to settle for 2 h, and imaged on a Zeiss Axio Observer fluorescence microscope before returning to the incubator. Cells were then stained with $1 \mu\text{M}$ calcein for 15 min daily, imaged, and returned to the incubator after imaging. Control experiments established that staining with $1 \mu\text{M}$ calcein did not affect cell growth rate or colony morphology. Cells were grown for 4 days, and then fixed and processed for RNA FISH experiments 16 h after the final calcein staining and imaging to allow for loss of calcein signal.

ChIP-seq. For ChIP-seq experiments, mESCs were grown in the absence of feeder cells and cross-linked when they were at $\sim 50\%$ confluence with 1% formaldehyde for 10 min at room temperature. Cross-linking was quenched by the addition of 125 mM glycine, and cells were then washed with PBS containing protease inhibitors, harvested by scraping and cell pellets were flash-frozen in liquid nitrogen. ChIP-seq was then performed as previously described⁶¹.

Antibodies used for chromatin immunoprecipitations were as follows: H3K4me3 (Millipore 07-473), H3K27me3 (Upstate 07-449), H3K36me3 (Abcam ab9050), H3K9me3 (Abcam ab8898), H3K27ac (Abcam ab4729) and RNAP2 (Covance MMS-128P).

To quantify ChIP-seq levels of H3K4me3, H3K27me3, H3K36me3, H3K9me3, H3K27ac and RNAP2 on a per-gene level, we calculated the number of reads which overlapped with the promoter region of each gene. Promoter regions were defined as the region from 1.5 kb upstream to 0.5 kb downstream of the transcription start site of each annotated gene in the NCBI37/mm9 RefSeq annotations.

Single-cell qRT-PCR. Single-cell qRT-PCR of mESCs was performed on a Fluidigm BioMark according to manufacturer's recommendations (full details available at <http://www.fluidigm.com>), with modifications as noted below. Briefly, cells were harvested from culture, stained using an antibody to SSEA1 (Clone MC-480, BD Biosciences), and individual mESCs were sorted into 8-well PCR strips or 96-well PCR plates on the basis of SSEA1 expression and gated such that the top $\sim 30\%$ of SSEA1-expressing cells were collected. The viability dye Sytox blue (Life Technologies) was used to exclude dead cells. Cells were sorted into 9 μl of reaction buffer containing a mixture of 96 different TaqMan Gene Expression Assays (Life Technologies) and reagents from the CellsDirect One-Step qRT-PCR kit (Life Technologies), and RT-PCR was performed with 22 cycles of specific target amplification (STA). Amplification products were diluted 1:5 with TE(8.0), and samples were

checked for the presence of amplified cDNA by duplex microlitre-scale qPCR using Gapdh-FAM and ActB-VIC TaqMan probes (Applied Biosystems). This step was performed to exclude wells that did not contain an actual cell or in which amplification failed, as initial experiments had found that a substantial fraction of wells did not contain amplification products, presumably due to inefficiencies in single-cell sorting. For wells showing successful amplification of both Gapdh and ActB (typically ~50–75% of samples on a plate), 5 µl of diluted samples were submitted for expression profiling, along with a set of population and negative control samples, on 96 × 96 gene expression chips on the Fluidigm BioMark. Ct values were determined with Fluidigm Real-time PCR analysis software, applying the auto (detectors) method and using a linear(derivative) baseline correction with a quality threshold of 0.65. When expression of a gene was not detected in a cell, its Ct value was assigned a value of 40 for the purposes of analysis. Data for the 84 probes that were common to all Fluidigm expression arrays performed are shown in Extended Data Fig. 6b and listed in Supplementary Table 4, along with TaqMan probe ID numbers. Amplification efficiencies for TaqMan probes were determined using standard curve analysis, and only probes with an efficiency between 0.8–1.25 per cycle were included in the study.

For the comparison of technical versus biological variability shown in Extended Data Fig. 6a, three pools of 100 or 10 cells or individual cells were sorted into separate wells of PCR strips and processed as described above, and pre-amplified cDNA from each pool or single cell was run in triplicate on a Fluidigm gene expression chip. To compare technical variability, the average of the three standard deviations of the technical replicates is indicated by the error bars in the *x*-direction in Extended Data Fig. 6a. To compare biological variability, the standard deviation of the average Ct value for the three separate pools or individual cells is indicated by the *y* error bars.

For the samples labelled 'Hoechst' in Fig. 3 and Extended Data Fig. 6, v6.5 mESCs cultured in serum + LIF were incubated with 10 µg ml⁻¹ Hoechst 33342 in serum + LIF growth media for 30 min at 37 °C before sorting, and single cells in the G₀/G₁ phase of the cell cycle were isolated for analysis.

To determine expression distributions changed significantly in response to perturbations, we compared the expression distribution of each gene (Ct value) in each condition to that in the v6.5 serum + LIF sample using the two-sample Kolmogorov–Smirnov test. *P* values were adjusted using the Holm method to correct for multiple tests.

Principal components analysis of qRT-PCR data and cell state classification.

The principal components analysis of single-cell qRT-PCR data (Fig. 3c) was performed by applying the *prcomp* R function to the complete data matrix of 1,144 individual cells and 81 genes. Genes contributing the most (with absolute loadings >0.18) to the first three principal components distinguishing mESC perturbations (excluding MEFs) are displayed in Fig. 3f. To identify distinct transcriptional states of the qPCR data, we modelled the first three principal components as multivariate normal mixtures, allowing the model selection of step of mClust to test 1 to 9 components and all model geometries available. We used the model parameters that maximized the Bayesian Information Criteria, resulting in a model of five components (that is, clusters or groups of transcriptional states), for further analysis (Extended Data Fig. 6). The resulting model is a mixture density of the form:

$$f(x) = \sum_{i=1}^C P_i f_i(x)$$

where *i* = component, *P_i* = the probability that the observation *x* comes from the component (cluster or group) *i*, and *f_i* is the probability density of component *i*. This model can be used to compute the posterior probability that a given cell comes from one of the *C* components using Bayes theorem:

$$P(x \in \text{cluster } A) = \frac{P_A f_A(x)}{\sum_{i=1}^C P_i f_i(x)}$$

Cells were classified based on the component that maximized this posterior probability. However, if the maximum posterior probability did not exceed 0.50, then the cell was not assigned a state (unclassified in Fig. 3e).

Quantitative immunofluorescence. For quantitative immunofluorescence experiments, mESCs were grown in the absence of feeder cells on Matrigel (BD)-coated 24-well glass bottom plates (MatTek, No. 1.5 glass, 13 mm diameter) and cross-linked with 4% paraformaldehyde (Alfa Aesar, catalogue no. 43368) for 15 min at room temperature. Immunostaining was performed according to standard protocols⁶². Briefly, cells were washed in PBS, permeabilized with 0.2% Triton X-100 plus either 1% normal goat serum (NGS, Life Technologies) or bovine serum albumin (BSA, Sigma) if goat primary antibodies were used, washed in PBS + NGS or BSA, and incubated in primary antibody overnight at 4 °C (see below for antibodies and concentrations used). The next day, samples were washed with PBS + NGS/BSA and incubated with 1:200 secondary antibody for 1 h at room temperature (see below for list of secondary antibodies used). Samples were then washed with

PBS, treated with 1 µg ml⁻¹ DAPI (Life Technologies) for 15 min at room temperature, washed and treated with ProLong Gold antifade reagent (Life Technologies), then covered with a cover slip and allowed to cure overnight.

Images were collected on a Leica SP5 X MP inverted confocal microscope out-fitted with a tunable white light laser at 20× magnification (0.7 NA, dry) in *x*, *y* and *z*. For the data shown in Fig. 2e and Extended Data Fig. 5b, c, images were acquired using standard detectors, and voltage and gain settings were maintained at a constant level for each set of images acquired for each pair of transcription factors. For the data shown in Fig. 4e and Extended Data Fig. 5d, images were acquired using HyD detectors in photon counting mode. *z*-intervals were spaced 1 µm apart, and 10 slices were taken centred around the area of maximum intensity for each region imaged. Four-channel imaging was performed to obtain brightfield, DAPI (ex 405 nm), TF 1 (ex 564 nm) and TF 2 (ex 647 nm) images.

Relative protein levels within individual cells were quantified with CellProfiler⁶². Two-dimensional maximum projections of confocal *z*-stacks were input into CellProfiler, and cells were identified on the basis of OCT4 staining to exclude differentiated cells from the analysis. Performing the analysis by identifying cell nuclei on the basis of DAPI staining gave qualitatively similar results.

Antibodies used for quantitative immunofluorescence were as follows. Primary antibodies: OCT4, sc-5279 mouse monoclonal (1:100, Santa Cruz) or ab19857 rabbit polyclonal (1:100, Abcam); NANOG, ab80892 rabbit polyclonal (1:100, Abcam); ESRB, PP-H6705-00 mouse monoclonal (1:100, R&D Systems); NR0B1, sc-841 rabbit polyclonal (1:100, Santa Cruz). Secondary antibodies: goat anti-mouse Cy5 (1:200, Life Technologies); goat anti-rabbit Alexa 568 (1:200, Life Technologies).

miRNA expression analyses. Cells used in the miRNA expression profiling experiment described in Fig. 5a were cultured under the conditions described above, and RNA was extracted and purified according to manufacturer's instructions using a Qiagen miRNeasy kit. miRNAs were then quantified according to manufacturer's protocols on a NanoString nCounter using the NanoString mouse miRNA Assay Kit, starting with 100 ng of purified RNA per sample. Data were normalized by dividing by the geometric mean of four housekeeping mRNAs (*Actb*, *B2m*, *Gapdh* and *Rpl19*). The NanoString mouse miRNA probe set includes eight negative control probes. Analysis of miRNAs was limited to those 427 probes with normalized counts that exceed the maximum negative control value in at least one of the six conditions profiled. Differentially expressed miRNAs were defined as those with |log₂(fold change)| > 2.

qPCR validation of miRNA expression shown in Extended Data Fig. 9 was performed on a Fluidigm BioMark using a 48 × 48 expression array. Starting with purified mRNA from cell samples (300 ng total RNA per sample), miRNAs were reverse transcribed in the presence of a pool of TaqMan miRNA RT primers according to manufacturer's instructions using the TaqMan MicroRNA Reverse Transcription kit (Applied Biosystems) in a 15 µl reaction volume. 2.5 µl of the reverse transcription reaction was then input into a 25 µl pre-amplification reaction performed using TaqMan PreAmp Master Mix and the corresponding pool of TaqMan miRNA PCR primers, and 14 cycles of pre-amplification were performed according to Fluidigm protocols. This amplification product was diluted 1:10 with TE(8.0), and 5 µl was submitted for expression profiling on a Fluidigm BioMark. Ct values were determined with Fluidigm Real-time PCR Analysis Software, applying the auto(detectors) method and using a linear(derivative) baseline correction with a quality threshold of 0.65. Expression differences were determined using the ΔΔCt method, normalizing to the geometric mean of four reference small RNAs (snoRNA 142, snoRNA 202, snoRNA 234 and U6 RNA) and comparing to v6.5 mESCs cultured in serum + LIF as the reference sample. Error bars represent the standard deviations of three technical replicate PCR reactions.

Identification of predicted and experimentally determined miRNA target gene sets.

Predicted targets of ESCC miRNAs, the let-7 family of miRNAs, and miR-152 were determined using the programs TargetScan⁶³ and miRmap⁶⁴. For TargetScan, we considered as predicted targets of these miRNA families the set of all predicted target genes in mouse, irrespective of site conservation. miRmap considers conservation in its prediction algorithm. The *P* values for changes in expression of predicted let-7 and miR-152 targets in the section 'miRNA balance and cell state' were calculated using miRmap predicted targets, while the plots shown in Fig. 5b show expression of TargetScan predicted target genes. Selected experimentally validated targets of let-7 or miR-152 that are not predicted by TargetScan (*Myc* for let-7, *Dnmt3b* for miR-152) are indicated in yellow in Fig. 5b.

Self-renewal assays. For the clonal self-renewal assays described in Fig. 5, cells were cultured as described above, plated at a density of 200 cells per well of a matrigel-coated Nunc tissue culture plate, and allowed to form colonies over the course of 5 days. Colonies were then stained using the Vector Red Alkaline Phosphatase Substrate Staining kit (Vector Labs) using 200 mM Tris-HCl (8.5) plus 150 mM NaCl as the staining buffer, and fixed with 4% paraformaldehyde for 10 min at room temperature. Colonies were imaged on an Axio Zoom V16 (Zeiss) and assigned as showing either solid or mixed alkaline phosphatase staining by visual inspection.

We observe a strong correlation between uniformity of alkaline phosphatase staining within individual colonies and the uniformity of Nanog protein expression they exhibit, and therefore interpret this assay as reflective of the self-renewal capacity of individual mESCs.

miRNA perturbation experiments. For the *Lin28a* knockdown experiments shown in Fig. 5d, iShLin28a cells were either treated with $1 \mu\text{g ml}^{-1}$ doxycycline or left untreated and maintained in serum + LIF media for 3 days. After 3 days of doxycycline treatment, cells were harvested and plated for self-renewal assays as described above, except that doxycycline was added to treated cells during self-renewal assays. For the forced let-7 experiments shown in Fig. 5e, iLet-7 cells⁴⁵ were treated with $10 \mu\text{g ml}^{-1}$ doxycycline or left untreated and cultured for 5 days, passaging once in between. By day 5 of doxycycline treatment, colony morphology showed a noticeable shift towards smaller, compact colonies. Cells were then harvested and plated for self-renewal or RNA FISH assays as described above, and doxycycline induction was continued during growth for these assays. For the miRNA inhibition experiments shown in Fig. 5g, mESCs grown in 2i + LIF media were transfected with miRNA inhibitors (hsa-let-7 miRCURY LNA microRNA Power family inhibitor, Product 460006; mmu-miR-152-3p miRCURY LNA Power microRNA inhibitor, Product 4101871-101; or miRCURY LNA microRNA Power Inhibitor Negative Control A, 199020-04, all from Exiqon) at a concentration of 20 nM each using Dharmafect 1 (Dharmacon) one day after plating at a concentration of 2,000 cells per well of a 24-well plate. Media was changed after 24 h, and cells were harvested 3 days after transfection and RNA was extracted using a Qiagen miRNeasy Micro kit for qPCR analysis. miRNA expression experiments shown in Extended Data Fig. 9 were performed essentially as described⁴⁴. *Dgcr8*^{-/-} mESCs maintained off of feeders were plated at a concentration of 5,000 cells per well of a 24-well plate, and were transfected the following day with miRIDIAN miRNA mimics (Dharmacon) at a concentration of 50 nM each using Dharmafect 1. Media was changed after 24 h, and cells were harvested 3 days after transfection and plated for self-renewal assays.

Polycomb target correlation with pluripotency factor expression. Polycomb targets were identified as described above ('Definition of Polycomb target genes for expression comparisons'). The proportion of Polycomb targets is the number of Polycomb targets that were detected ($\text{Ln}[\text{TPM}] > 0$) in each individual cell divided by the number of Polycomb targets that were detected in at least one cell profiled by RNA-seq (873). By this metric, any individual cell expressed between 6.0–21.5% of Polycomb target genes (Extended Data Fig. 8a). Pluripotency factor z-scores were calculated for each condition as:

$$z - \text{score}(\text{sample}_s, \text{gene}_g, \text{condition}_c) = \frac{\text{expression}_{s,g,c} - \mu(\text{expression})_{g,c}}{\sigma(\text{expression})_{g,c}}.$$

Pearson correlation coefficients were calculated on pluripotency factor z-scores and proportion of Polycomb targets expressed for each condition (Extended Data Fig. 8b).

Statistics. Below, we describe the statistical tests and/or sample sizes of all data comparisons shown in the main and Extended Data figures, or mentioned in the main text. Holm's method corrected *P*-values are reported⁴⁶ unless otherwise stated. No statistical method was used to predetermine sample size.

The *P* values for differences in H3K27me3 levels between detected and non-detected Polycomb target genes in 'The landscape of gene expression variability in PSCs' section were calculated using the Student's *t*-test. *P* values for the significance of lower expression of let-7 and miR-152 target genes and higher expression of ESCC miRNA targets in 2i + LIF in the section 'miRNA balance and cell state' were calculated using the Mann-Whitney test. Calculation of the correlation in expression between predicted let-7 and miR-152 target genes is described in the 'Single-cell correlation between expression of predicted let-7 or miR-152 target genes' section of Supplementary Information. *P* values for significance of differences between high and low states shown in Fig. 2g were calculated using the Kolmogorov-Smirnov test.

For Fig. 1b, the number of colonies and cells examined was 318 cells from 5 colonies (*Oct4*), 1,226 from 10 colonies (*Esrrb*), and 551 cells from 5 colonies (*Neurod1*). For Fig. 1c, COV enrichment of biological pathway annotations. COV-based enrichment analysis was performed by testing whether genes in each Gene Ontology Biological Process⁶⁶ and Reactome⁶⁷ annotation tended to have higher or lower COVs in serum + LIF cells than would be expected by chance given the distribution of μ in the gene set.

For Fig. 2b, cell colony numbers: 258 cells for the colony on the left in the high expression state, 70 cells for the colony in the middle in the low expression state, and 98 cells for the colony on the right in the mixed expression state. These three colonies are part of the 10 colonies examined for the overall *Esrrb* population distribution in serum + LIF shown in Figs 1b and 2c. Fig. 2c, *Oct4* and *Dazl*: 5 colonies, 318 cells; *Sox2* and *Neurod1*: 5 colonies, 551 cells; *Nanog* and *Otx2*: 5 colonies, 151 cells; *Esrrb*: 10 colonies, 1,226 cells; *Lin28a*: 5 colonies, 472 cells; *Nr0b1*: 8 colonies,

849 cells; *Myc*, 8 colonies, 867 cells; *Olig2*, 10 colonies, 1,097 cells; *Nkx2-9*: 5 colonies, 524 cells; and *Pax6*: 5 colonies, 417 cells. *Oct4* and *Dazl*, *Sox2* and *Neurod1*, and *Nanog* and *Otx2* were each imaged by two-colour RNA FISH in the same set of cells. Average expression values and standard deviations for these data are shown in Extended Data Fig. 3a. *Esrrb* and *Neurod1* individual colony data comprise the population distributions shown in Fig. 1b.

For Fig. 4d, serum + LIF: 1,226 cells from 10 colonies (same data as in Figs 1b and 2c); 2i + LIF: 557 cells from 5 colonies; *Dgcr8*^{-/-}: 187 cells from 5 colonies. Fig. 4e, serum + LIF: 901 cells; 2i + LIF: 1,648 cells; *Dgcr8*^{-/-}: 248 cells.

For Fig. 5a, miRNAs with $|\log_2(\text{fold change})| > 2$ (mESC in 2i + LIF versus FBS + LIF) are shown.

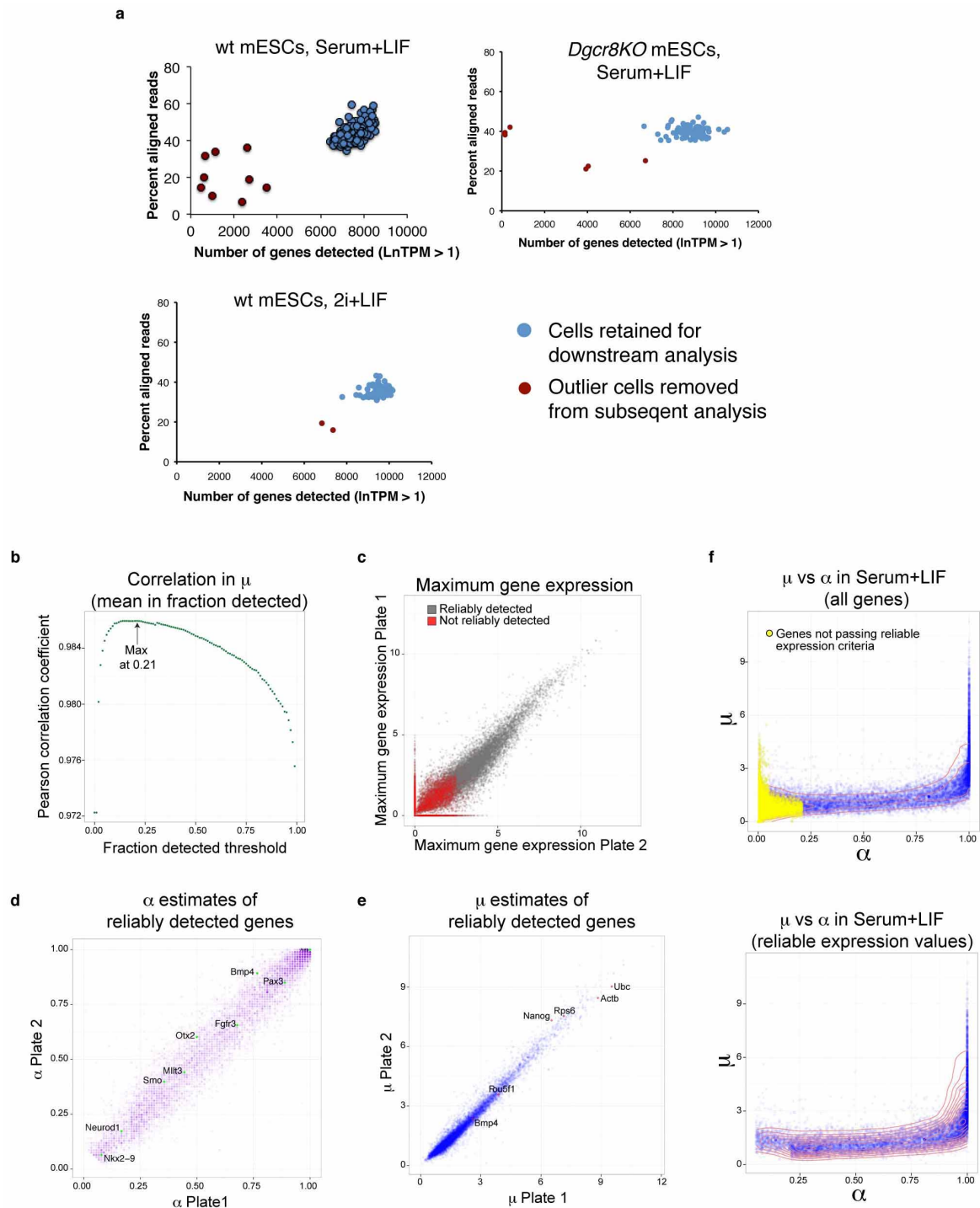
For Extended Data Fig. 6b, number of cells in each condition: v6.5 on MEFs (75), v6.5 FF (83), E14 FF (77), v6.5 2i + LIF (77), E14 2i + LIF (87), *Dgcr8*^{-/-} (81), iPSCs (57), Hoechst (84), *Dnmt1*^{-/-} *Dnmt3a*^{-/-} *Dnmt3b*^{-/-} (34), *Eed*^{-/-} (54), BIX (30), TSA (30), VPA (30), Mbd3 (40), EpiSCs (41). Extended Data Fig. 6c, changes in the distributions of gene expression (by qPCR) were tested by two-sample Kolmogorov-Smirnov on the Ct values in each perturbation or condition versus the v6.5 serum + LIF sample, and *P* values were adjusted using the Holm's method to correct for multiple tests.

For Extended Data Fig. 5b, NANOG and OCT4 in serum + LIF: 2,469 cells. Extended Data Fig. 5c, NR0B1 and OCT4 in serum + LIF: 2,917 cells.

For Extended Data Fig. 7e, *Oct4*: serum + LIF (5 colonies, 318 cells (same data as in Figs 1b and 2c)), 2i + LIF (5 colonies, 485 cells), *Dgcr8*^{-/-} (5 colonies, 175 cells). *Nanog*: serum + LIF (5 colonies, 151 cells (same data as in Fig. 2c)), 2i + LIF (5 colonies, 398 cells), *Dgcr8*^{-/-} (5 colonies, 145 cells). *Nr0b1*: serum + LIF (8 colonies, 849 cells (same data as in Fig. 2c)), 2i + LIF (5 colonies, 429 cells), *Dgcr8*^{-/-} (5 colonies, 283 cells). *Otx2*: serum + LIF (5 colonies, 151 cells (same data as in Fig. 2c)), 2i + LIF (5 colonies, 398 cells), *Dgcr8*^{-/-} (5 colonies, 144 cells). *Nanog* and *Otx2* were imaged together in the same set of cells using two colour RNA FISH. Extended Data Fig. 7g, average expression values in serum + LIF are the same shown in Extended Data Fig. 3a. For 2i + LIF, average expression values are given for *Oct4*, *Nanog*, *Nr0b1* and *Otx2* for the distributions shown in Extended Data Fig. 7e (above). For the remaining 2i + LIF genes, numbers of cells are as follows: *Myc* and *Isl1*: 6 colonies, 785 cells; *Lin28a* and *Olig2*: 5 colonies, 354 cells; *Esrrb*: 10 colonies, 903 cells; *Nkx2-9*: 5 colonies, 345 cells; *Sox2* and *Neurod1*: 5 colonies, 649 cells; *Pax6*: 5 colonies, 429 cells; and *Dazl*: 5 colonies, 485 cells. *Oct4* and *Dazl*, *Sox2* and *Neurod1*, *Myc* and *Isl1*, *Lin28a* and *Olig2*, and *Nanog* and *Otx2* were each imaged by two-colour RNA FISH in the same set of cells. Extended Data Fig. 7h, serum + LIF distributions are the same shown in Extended Data Fig. 5b, c. NANOG and OCT4 in 2i + LIF: 518 cells. NR0B1 and OCT4 in 2i + LIF: 827 cells.

49. Sommer, C. A. et al. Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells* **27**, 543–549 (2009).
50. Park, I. H., Lerou, P. H., Zhao, R., Huo, H. & Daley, G. Q. Generation of human-induced pluripotent stem cells. *Nature Protocols* **3**, 1180–1186 (2008).
51. Dow, L. E. et al. A pipeline for the generation of shRNA transgenic mice. *Nature Protocols* **7**, 374–393 (2012).
52. Lee, S. H., Lumelsky, N., Studer, L., Auerbach, J. M. & McKay, R. D. Efficient generation of midbrain and hindbrain neurons from mouse embryonic stem cells. *Nature Biotechnol.* **18**, 675–679 (2000).
53. Okabe, S., Forsberg-Nilsson, K., Spiro, A. C., Segal, M. & McKay, R. D. Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. *Mech. Dev.* **59**, 89–102 (1996).
54. Ramsköld, D. et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnol.* **30**, 777–782 (2012).
55. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
56. Fujita, P. A. et al. The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* **39**, D876–D882 (2010).
57. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
58. Patro, R., Mount, S. M. & Kingsford, C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nature Biotechnol.* **32**, 462–464 (2014).
59. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods* **5**, 877–879 (2008).
60. Carpenter, A. E. et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
61. Garber, M. et al. A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol. Cell* **47**, 810–822 (2012).
62. Spector, D. L. & Smith, H. C. Redistribution of U-snrNPs during mitosis. *Exp. Cell Res.* **163**, 87–94 (1986).
63. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).

64. Vejnar, C. E. & Zdobnov, E. M. MiRmap: comprehensive prediction of microRNA target repression strength. *Nucleic Acids Res.* **40**, 11673–11683 (2012).
65. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
66. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature Genet.* **25**, 25–29 (2000).
67. Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **37**, 619–622 (2009).
68. Fraley, C. & Raftery, A. E. Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.* **458**, 611–631 (2002).
69. Cahan, P. *et al.* CellNet: network biology applied to stem cell engineering. *Cell* **158**, 903–915 (2014).

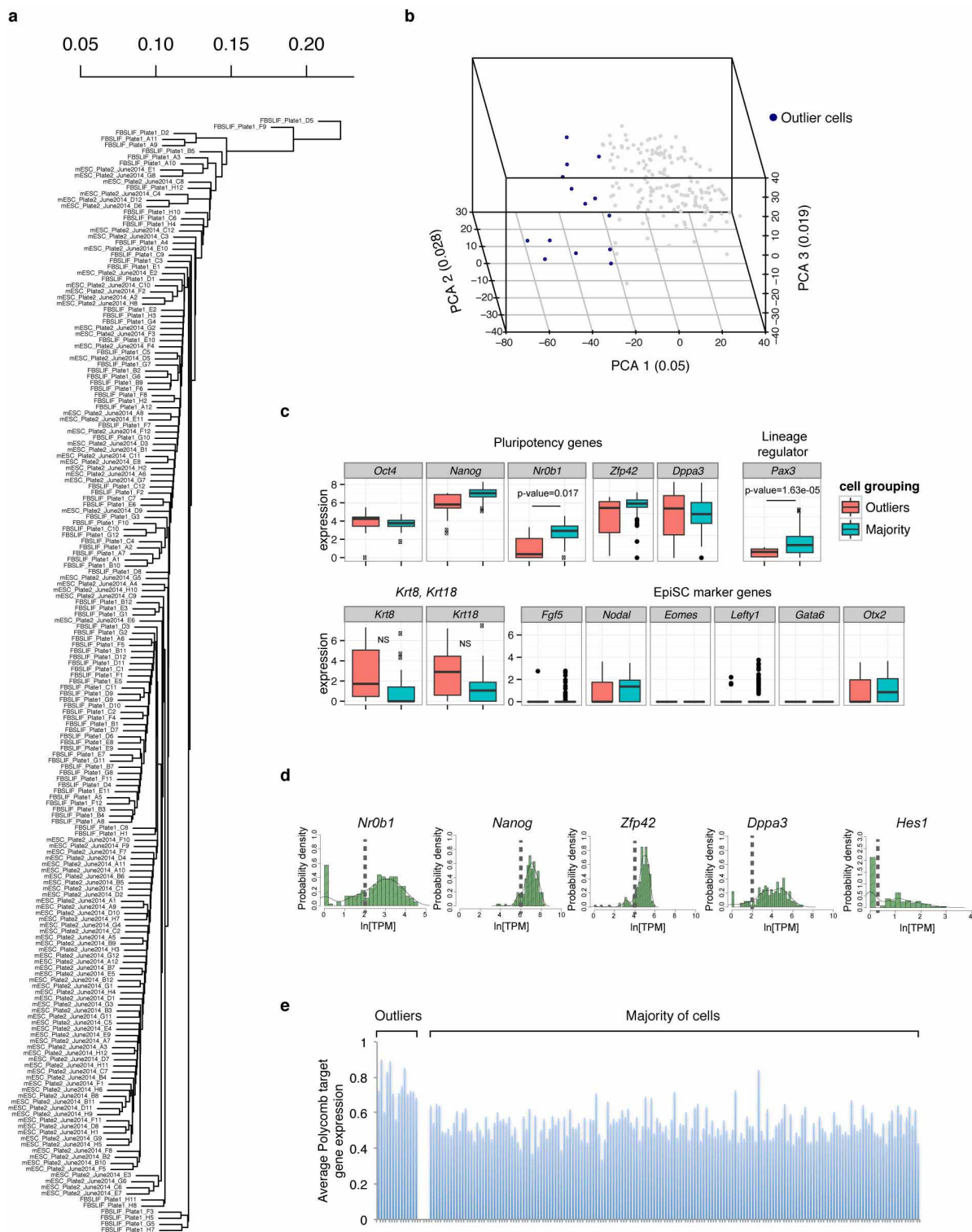


Extended Data Figure 1 | Quality control of single-cell RNA-seq data. **a**, A combination of read alignment rate (y axis) and number of genes detected ($\ln[\text{TPM}] > 1$) (x axis) was used to identify outlier cells (red circles) to remove from subsequent analysis, leaving 183 single mESCs cultured in serum + LIF, 94 mESCs cultured in 2i + LIF, and 84 *Dgcr8*^{-/-} mESCs cultured in serum + LIF that were analysed by single-cell RNA-seq in this study (blue circles).

b, Correlation in mean expression in detected cells (μ) between replicate serum + LIF plates across a range of α (fraction of cells a gene was detected in) thresholds. μ was calculated separately for each plate. Correlations in μ were calculated after limiting genes to those with α thresholds exceeding the specified

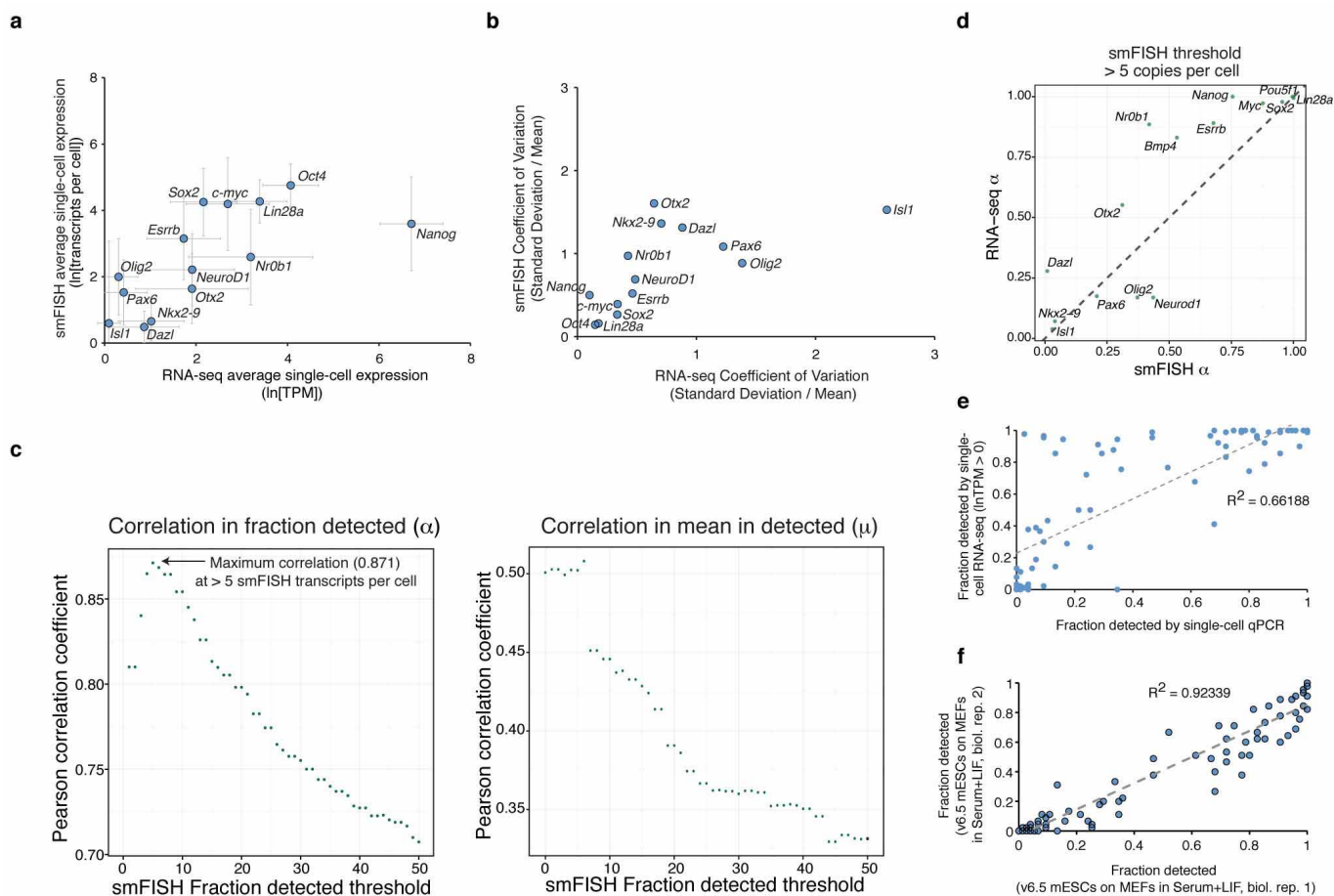
threshold on the x axis. **c**, Maximum gene expression in replicate plates. Genes not reliably detected as defined in the text are coloured red.

d, Comparison of α estimates in replicate plates, limited to reliably detected genes. Selected lineage regulators are denoted. **e**, Comparison of μ estimates in replicate plates, limited to reliably detected genes. Selected housekeeping, pluripotency and signalling genes are denoted. **f**, Top: relationship between estimates of α (x axis) and μ (y axis) of all genes based on both plates of mESC in FBS + LIF. Undetected genes are coloured yellow. Bottom: same as above except only showing reliably detected genes, and overlaid with density contour (red lines).



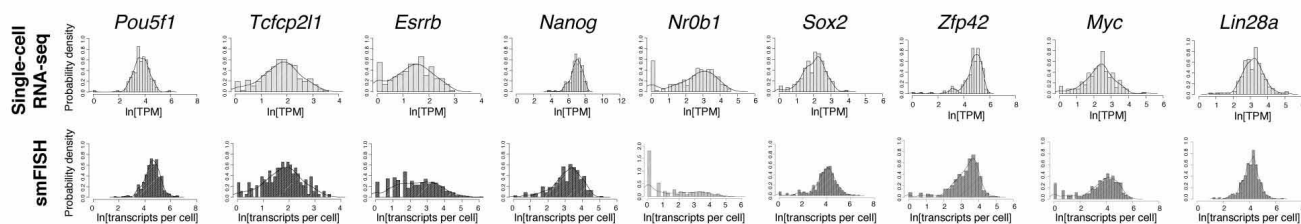
Extended Data Figure 2 | Examination of mESCs cultured in serum+LIF for the presence of distinct subpopulations. **a**, Hierarchical clustering dendrogram of single-cell RNA-seq data for 183 mESCs cultured in serum+LIF. **b**, Principal component analysis of the 183 mESCs cultured in serum+LIF. Points coloured blue are those with PCA1 values < -25 , which are classified as outlier cells. **c**, Box plots of expression of selected pluripotency regulators and the lineage regulator *Pax3* (top), and genes associated with EpiSCs (bottom). 'Normal' indicates the majority of cells coloured as grey dots

in Extended Data Fig. 2b; 'Outliers' indicates the distinct set of 14 cells coloured blue in Extended Data Fig. 2b. *P* values for statistically significant differences are shown. **d**, Histograms showing the expression distributions of pluripotency regulators previously found to fluctuate within mESC populations. Cutoffs to divide expression into high and low states to test for enrichment within outlier cells are indicated by dashed lines. **e**, Average expression of Polycomb target genes within outlier cells (left), and the majority of the mESCs cultured in serum+LIF (right).

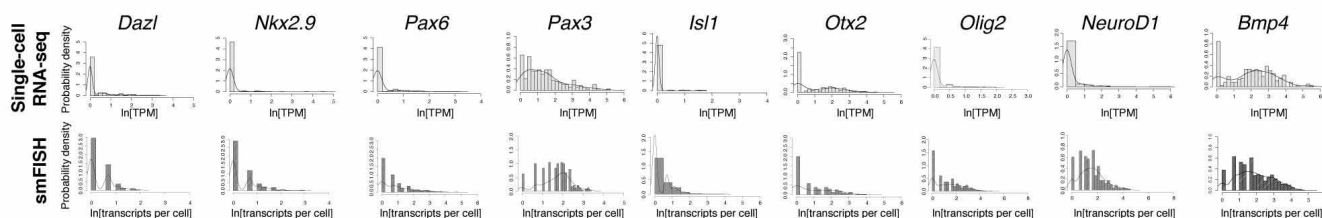


g

Pluripotency regulators



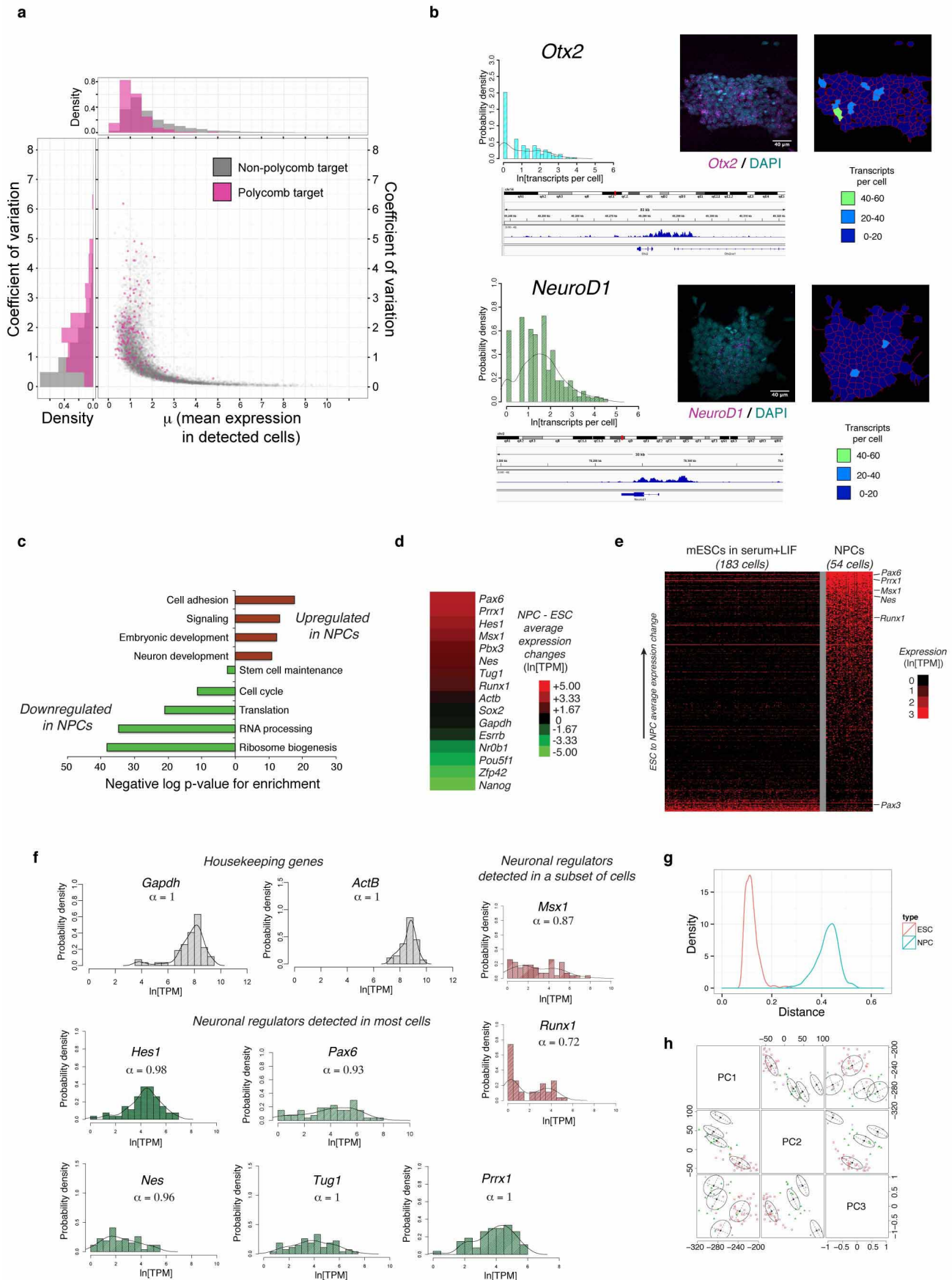
Lineage regulators / Polycomb targets



Extended Data Figure 3 | Correlation of single-cell expression

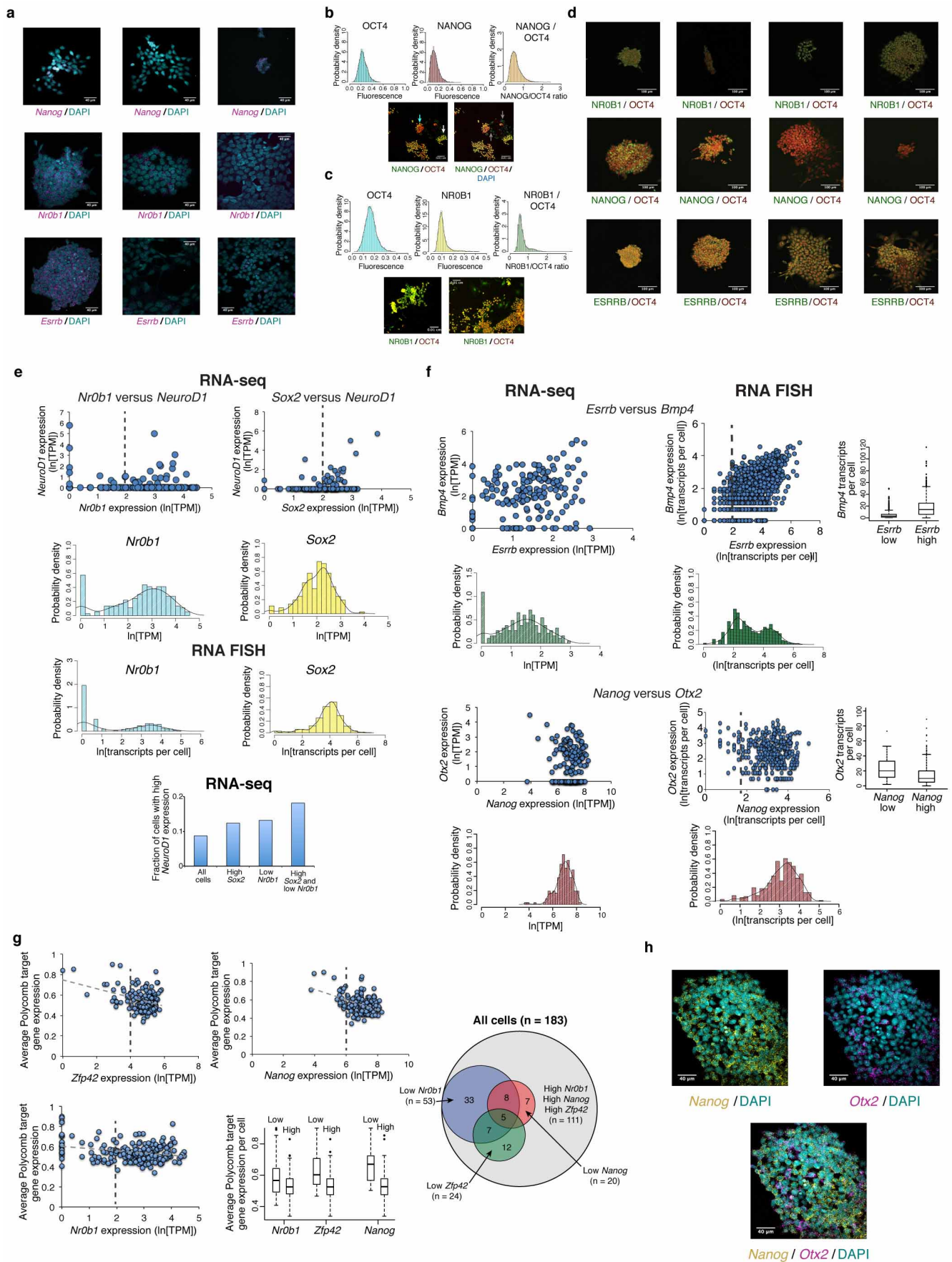
measurements between different technologies. **a**, Correlation between transcripts per million (TPM) measured by single-cell RNA-seq and transcripts per cell measured by single-molecule FISH for 14 selected genes in mESCs cultured in serum+LIF. Error bars represent standard deviations of measurements. **b**, Correlation between coefficients of variation measured by single-cell RNA-seq and FISH for the 14 genes shown in panel **a**. **c**, Correlation coefficients for α (fraction detected, left) or μ (mean expression in detected cells, right) between single-molecule FISH and single-cell RNA-seq are plotted as a function of varying the threshold level for detection by RNA FISH (x axis). An RNA FISH detection threshold of 10 indicates that genes expressed at <10 copies per cell would not be detected by RNA FISH. Correlation for α between RNA-seq and RNA FISH peaked at an RNA FISH detection threshold of >5 transcripts per cell, giving an estimated single-cell RNA-seq detection efficiency of $\sim 20\%$ (1 out of 5 transcripts detected, assuming single-molecule sensitivity for the RNA FISH method). **d**, Correlation in α between

single-cell RNA-seq and single-molecule FISH for 14 genes measured by both methods, assuming a single-molecule FISH detection threshold of >5 transcripts per cell. Dashed line shows linear fit to the data. The fraction of cells a gene is detected in shows good agreement between the two methods when taking the sensitivity of the RNA-seq into account. **e**, Comparison of the fraction of mESCs cultured in serum+LIF a gene was detected in by single-cell qPCR (x axis) or single-cell RNA-seq (y axis). Single-cell RNA-seq showed greater sensitivity overall as compared to single-cell qPCR, but a set of genes was sporadically expressed as measured by both methods. Trend line indicates linear fit to the data. **f**, Correlations of fraction detected between independent biological replicates for 96 genes profiled by single-cell qPCR. Trend line shows linear fit to the data, and indicates that the fraction of cells a gene is detected in remains consistent across independent biological replicates. **g**, Comparison of expression distributions measured by single-cell RNA-seq (light grey) and single-molecule FISH (darker grey) for pluripotency regulators (top) and Polycomb target genes (bottom).



Extended Data Figure 4 | Expression of Polycomb target genes in ESCs and neural precursor cells (NPCs). **a**, Polycomb target genes show highly variable expression in mESCs. Relationship between μ (mean expression in detected cells, x axis) and coefficient of variation (standard deviation normalized by mean expression in all cells, y axis) is shown for Polycomb target genes (purple) and non-Polycomb target genes (grey). Distributions for μ and coefficients of variation for the two gene sets are shown above and to the left of the graph, respectively. Polycomb target genes show pronounced variability in expression, even when controlling for expression level. **b**, Expression of the neural regulators and Polycomb target genes *Otx2* (top) and *Neurod1* (bottom) measured by RNA FISH in mESCs cultured in serum+LIF. Overall distributions within the population (left) and representative colonies (right) are shown, along with gene tracks from the Broad Institute Integrated Genomics Viewer (IGV) showing ChIP-seq reads for H3K27me3 at the *Otx2* and *Neurod1* genes. **c**, Enriched gene ontology categories among genes significantly upregulated (red) or downregulated (green) in neural precursor cells as compared to embryonic stem cells. **d**, Expression changes of selected genes in neural precursor cells as compared to the embryonic stem cells

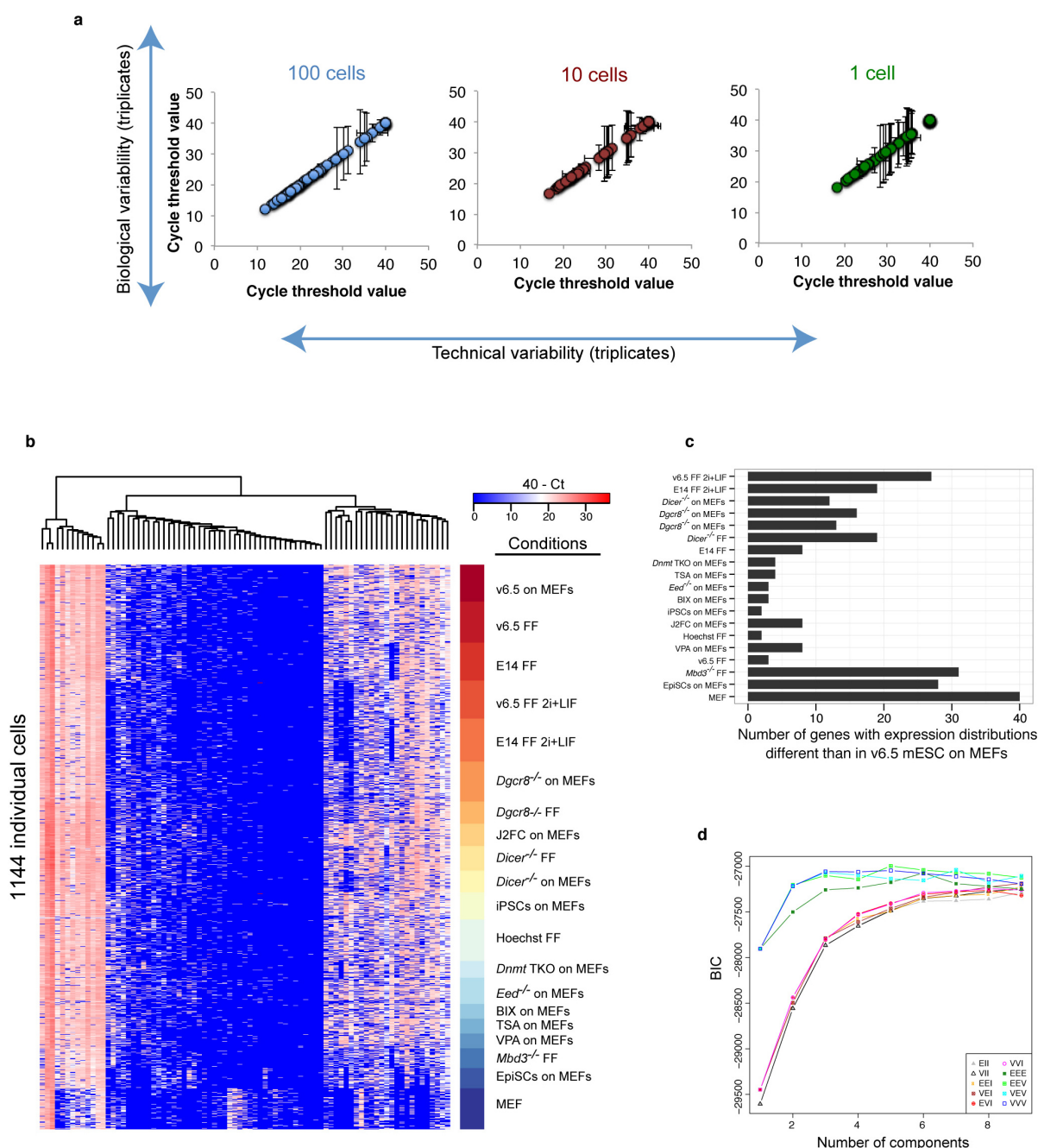
they were derived from. As expected, neural regulators and ES Polycomb target genes *Pax6*, *Prrx1*, *Hes1*, *Msx1*, *Pbx3*, *Nes* and *Runx1* were upregulated in NPCs, while the pluripotency regulators *Esrrb*, *Nr0b1*, *Pou5f1* (*Oct4*), *Zfp42* (*Rex1*) and *Nanog* were downregulated. Expression of the housekeeping genes *Gapdh* and *Actb*, and the pluripotency and neural regulator *Sox2*, were relatively unchanged between the two cell types. **e**, Expression comparison of ES Polycomb target genes that are detected in either mESCs or NPCs. Many Polycomb target genes that are neuronal regulators are detected in a higher fraction of NPCs than ESCs, while certain Polycomb targets such as *Pax3* (a regulator of musculoskeletal development) are detected in a smaller fraction of cells. Genes are ordered in ascending order of ESC to NPC average expression change. **f**, Histograms showing distributions of expression levels for selected housekeeping genes (grey) and neuronal regulators (green and red) in NPCs. The neuronal regulators *Msx1* and *Runx1* (red) show bimodal expression in NPCs. **g**, Distance distributions within ESC (red) and NPC (blue) populations. NPCs show more population heterogeneity than ESCs. **h**, State classification based on principal component analysis of single-cell RNA-seq of NPCs. Four distinct states are identified.



Extended Data Figure 5 | Fluctuations in pluripotency and lineage

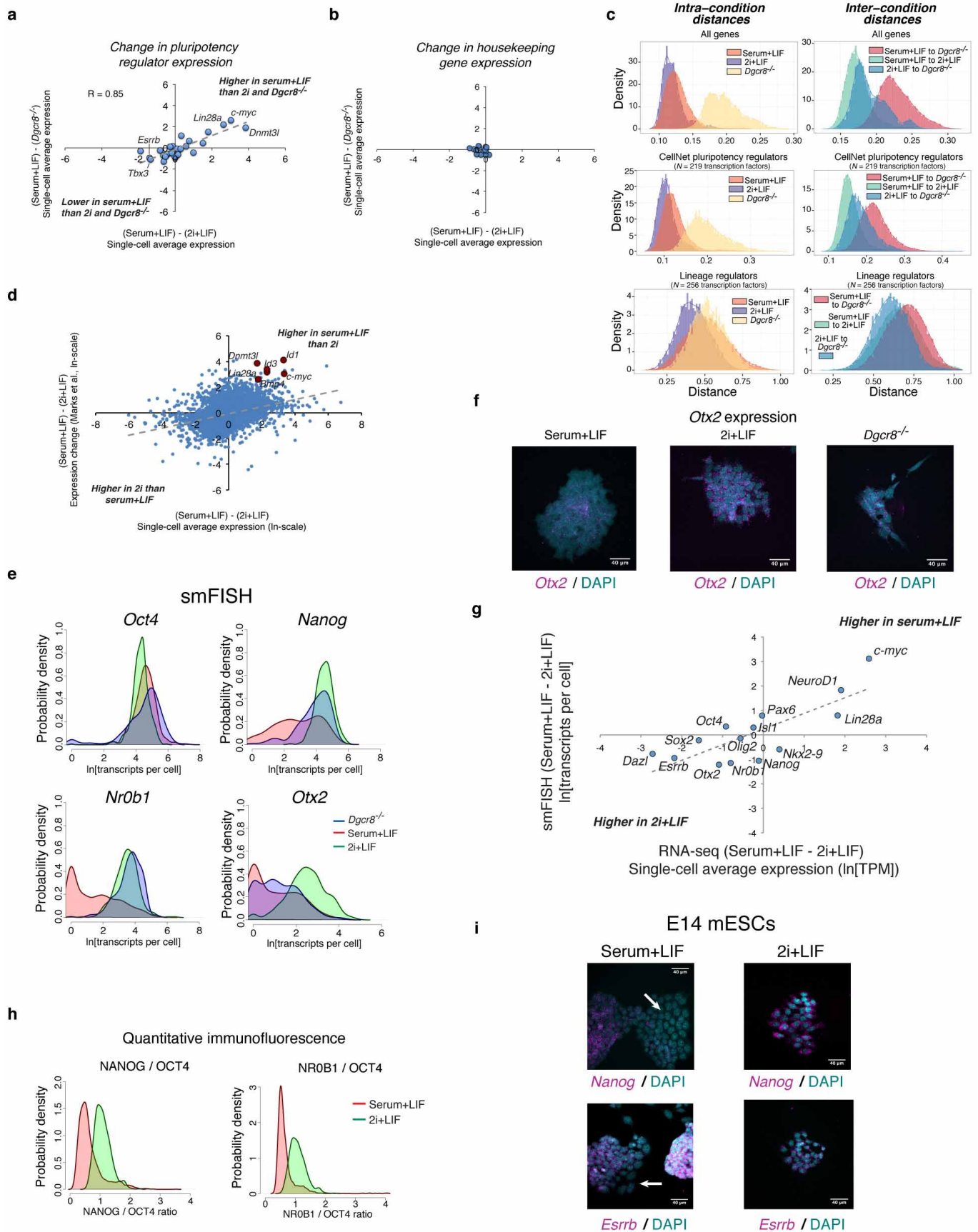
regulator expression. **a**, RNA FISH images showing expression of *Nanog* (top), *Nr0b1* (middle) and *Esrrb* (bottom) in individual colonies or regions of cells. **b, c**, Histograms show distributions of fluorescence intensities within individual cells from quantitative immunofluorescence of OCT4 and (b) NANOG, or (c) NR0B1, along with NANOG/OCT4 or NR0B1/OCT4 ratios as indicated. For NANOG/OCT4 images (b), NANOG staining is coloured green while OCT4 staining is coloured red. In the panel on the left a cluster of low NANOG cells is indicated by a blue arrow, while a cluster of high NANOG cells is indicated by a white arrow. In the panel on the right, the same image is shown with DAPI staining coloured blue, and groups of OCT4 negative/NANOG negative differentiated cells that may have arisen from the low NANOG cells are indicated with grey arrows. For NR0B1/OCT4 images (c), NR0B1 is coloured green while OCT4 staining is coloured red. A relatively high NR0B1 colony is shown in the panel on the left, while a region of low NR0B1 cells is displayed in the panel on the right. **d**, Quantitative immunofluorescence images showing expression of OCT4 (red) and NR0B1 (green, top row), NANOG (green, middle row), or ESRRB (green, bottom row) within individual colonies of mESCs grown in serum+LIF. OCT4 is used as an internal reference as it shows relatively invariant expression within mESCs. **e**, Single-cell RNA-seq relationships for gene pairs shown in Fig. 2g.

Distributions of gene expression from RNA-seq and RNA FISH experiments are shown. Dashed lines indicate divisions between high and low states for box plot shown in Fig. 2g. Single-cell RNA-seq data show that the subset of cells in both a high *Sox2* and low *Nr0b1* state show an increased probability of expressing *Neurod1* ($\ln[\text{TPM}] > 1$) as compared to all cells (bottom). **f**, RNA-seq (left) and RNA FISH (right) correlations between the pluripotency regulator *Esrrb* and the signalling factor *Bmp4* (top), and the pluripotency regulator *Nanog* and the neural regulator *Otx2* (bottom). Dashed lines indicate divisions between high and low states for the box plots shown on the right. **g**, Correlations from single-cell RNA-seq data between average Polycomb target gene expression and *Zfp42* (*Rex1*), *Nanog* and *Nr0b1*. Dashed lines indicate divisions between high and low states for the box plot comparing Polycomb target expression with expression of the three regulators, which are all negatively associated with expression of Polycomb targets. The Venn diagram shows coupling between high and low states of the three regulators. Low *Nr0b1* cells are more likely to be in a low *Zfp42* and/or low *Nanog* state as compared to high *Nr0b1* cells, suggesting that *Nr0b1* functions to maintain *Zfp42* and *Nanog* expression and repress Polycomb target genes. **h**, RNA FISH images of an ESC colony hybridized with probes against *Nanog* (yellow) and *Otx2* (magenta), showing inverse relationship between *Nanog* and *Otx2* expression.



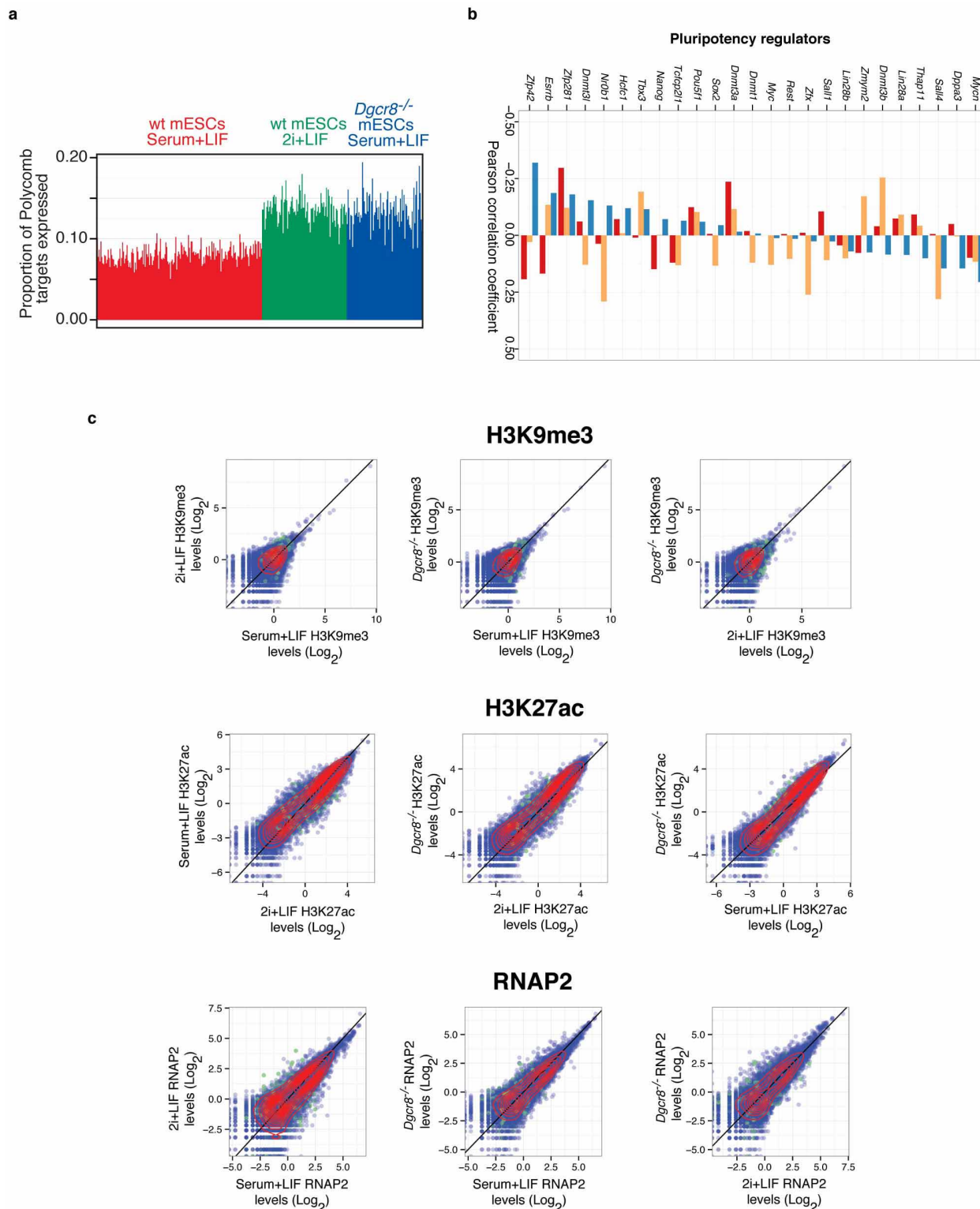
Extended Data Figure 6 | Single-cell qPCR of mESCs exposed to chemical and genetic perturbations. **a**, Shown are the average Ct values and standard deviations for technical triplicates (error bars on x axis) or biological triplicates (error bars on y axis) across 96 genes for pools of 100 or 10 cells or single sorted cells. Cells were sorted into PCR strips containing RT-PCR reagents and primer pools, reverse transcription and pre-amplification was performed, and cDNA was quantified on a Fluidigm BioMark PCR system. **b**, Heat map of single-cell qPCR data for 84 genes examined across 19 different PSC perturbations and in MEFs ($n = 1,144$ single cells). Unsupervised hierarchical clustering grouped genes into three clusters: bimodally expressed genes (right group), ubiquitously expressed genes (left group), and sporadically

expressed genes (middle group). **c**, Numbers of genes showing significant changes in expression distributions as compared to the reference conditions of v6.5 mESCs cultured in serum+LIF on MEFs. Significance of changes was determined by the two-sample Kolmogorov-Smirnov test, correcting P values for multiple tests using Holm's method. **d**, Selection of the state classification model that maximizes the Bayesian Information Criteria (BIC, y axis). Mclust was used to generate multivariate Gaussian mixture models of the first three principal components of the Fluidigm qPCR-based expression values of individual mESCs. The models vary in the number of components (one to nine) and the following geometric characteristics: volume, shape and orientation as described⁶⁸. The best model was used to classify cells into states.



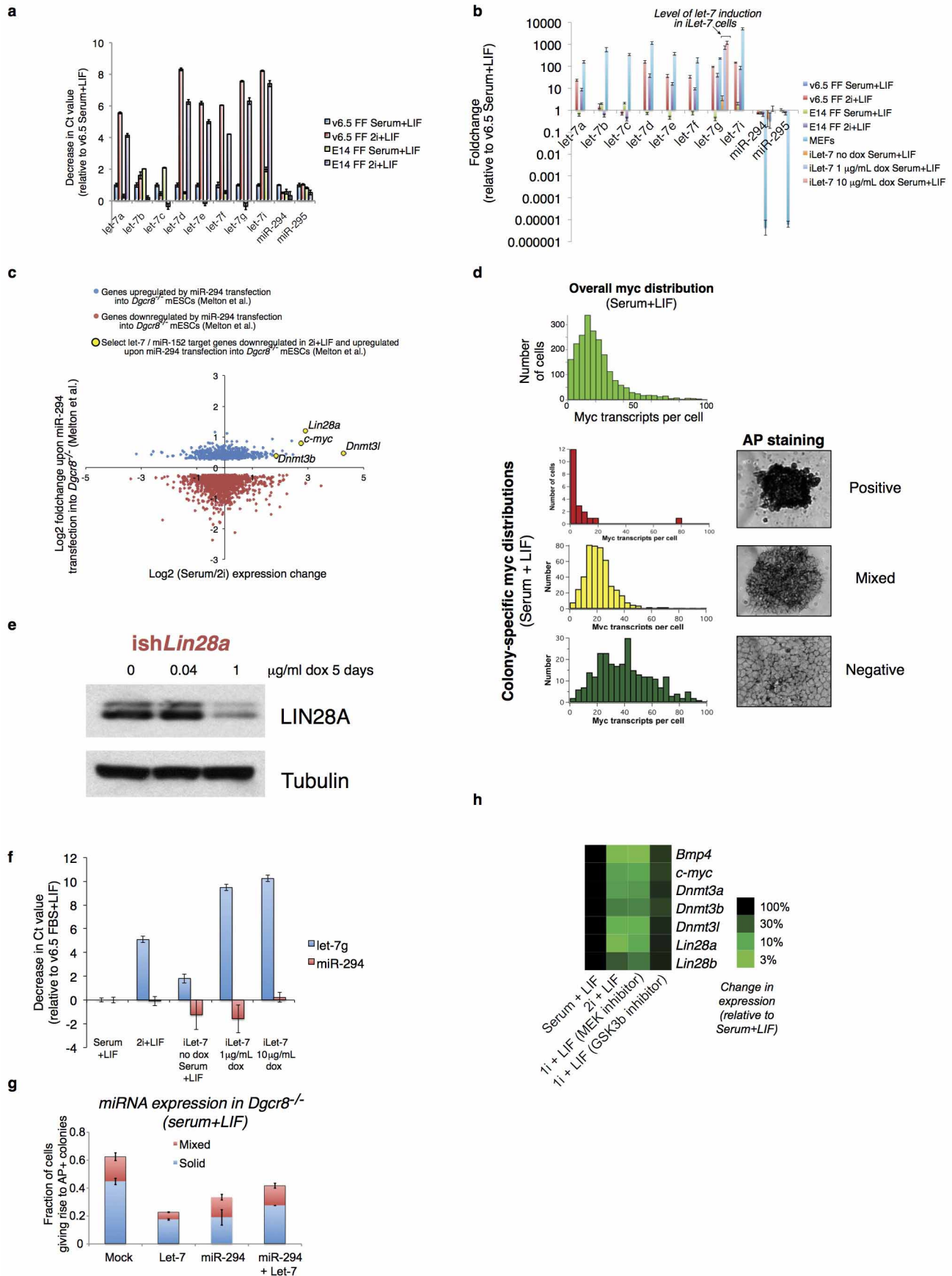
Extended Data Figure 7 | Gene expression changes in mESCs upon culture in 2i or removal of mature miRNAs. **a**, Changes in expression of the pluripotency regulators shown in Fig. 4c going from serum+LIF to 2i+LIF culture (*x* axis), or between wild-type and *Dgcr8*^{-/-} cells cultured in serum+LIF (*y* axis), as measured by single-cell RNA-seq. Selected genes are highlighted. **b**, Changes in expression of 18 commonly used housekeeping genes (*Actb*, *Aip*, *Cxcl1*, *Gapdh*, *Gusb*, *Hmbs*, *Hprt*, *Ipo8*, *Mrpl48*, *Mtp1*, *Pgk1*, *Ppia*, *Rpl13a*, *Rplp2*, *Rps6*, *Tbp*, *Ubc* and *Ywhaz*) between the same conditions as in panel **a**. **c**, Intra-condition (left) and inter-condition (right) distances between individual cells based on single-cell RNA-seq data for all genes (top), 219 transcription factors that regulate pluripotent cells as determined by CellNet⁶⁹ (middle), or lineage regulators, defined as the 256 previously determined Polycomb target genes in mESCs that are transcription factors (bottom). **d**, Comparison of single-cell average expression changes going from serum+LIF to 2i+LIF culture in the present study (*x* axis) against population-level expression changes between mESCs cultured in serum+LIF versus 2i+LIF measured by ref. 40 (*y* axis). Trend line from linear fit to the data are shown, and selected genes that show lower expression in 2i+LIF culture in both

studies are highlighted. **e**, Single-molecule FISH showing shifts in expression of *Oct4*, *Nanog*, *Nr0b1* and *Otx2* at the RNA level between wild-type mESCs in serum+LIF and 2i+LIF culture conditions and *Dgcr8*^{-/-} mESCs cultured in serum+LIF. **f**, Representative RNA FISH images showing expression of the Polycomb target gene and neural regulator *Otx2* in individual mESC colonies under the three conditions examined. **g**, Correlation between expression shifts between serum+LIF and 2i+LIF culture observed by single-cell RNA-seq (*x* axis) and RNA FISH (*y* axis) for the 14 genes shown in Extended Data Fig. 3a. Trendline indicates linear fit to the data. **h**, Quantitative immunofluorescence showing changes in NANOG/OCT4 (left) and NR0B1/OCT4 (right) ratios between serum+LIF and 2i+LIF culture. Serum+LIF data are the same shown in Extended Data Fig. 5. **i**, RNA FISH images of E14 mESC colonies cultured in serum+LIF (left) or 2i+LIF (right) media and probed for *Nanog* (top) or *Esrrb* (bottom) expression. Both *Nanog* and *Esrrb* show bimodal expression patterns in E14 mESCs grown in serum+LIF, and shift towards the high expression state in 2i+LIF culture. White arrows indicate regions of low *Nanog* or *Esrrb* expression in mESCs grown in serum+LIF.



Extended Data Figure 8 | Dependence of Polycomb target gene expression on culture conditions and miRNAs. **a**, Fraction of Polycomb target genes detected in wild-type mESCs cultured in serum+LIF (red) or 2i+LIF (green), or *Dgcr8*^{-/-} mESCs cultured in serum+LIF (blue). **b**, Correlation between Polycomb target gene expression and pluripotency regulator expression in different conditions (serum+LIF in blue, *Dgcr8*^{-/-} in orange, and 2i+LIF in red). Displayed are the Pearson correlation coefficients (PCC) between pluripotency-related regulator z-score and proportion of Polycomb targets detected, computed across all single cells. The z-score is defined as the number of standard deviations that a sample exceeds (z-score >0) or is less than (<0) the mean value. z-scores for pluripotency regulators were computed for each condition separately. A low PCC indicates that a lower factor expression

(relative to its mean in the condition) increases the likelihood that Polycomb targets will be detected as expressed (for example, *Zfp42* in FBS+Lif). **c**, Scatter plots comparing amount of H3K9me3 (top), H3K27ac (middle) and RNA polymerase II (bottom) at promoter regions in wild-type mESCs cultured in serum+LIF versus 2i+LIF conditions, wild-type mESCs versus *Dgcr8*^{-/-} mESCs cultured in serum+LIF, or *Dgcr8*^{-/-} mESCs cultured in serum+LIF versus wild-type mESCs cultured in 2i+LIF as indicated. ChIP-seq reads at gene promoters were median normalized for comparison, and Polycomb target genes are indicated in green. Unlike H3K27me3, levels of these three factors do not show a strong decrease at Polycomb target genes under 2i+LIF conditions and in *Dgcr8*^{-/-} mESCs (compare to Fig. 4f).



Extended Data Figure 9 | Perturbing miRNA balance and the *Myc/Lin28/let-7* axis. **a**, Purified RNA from v6.5 mESCs and E14 mESCs cultured in either serum+LIF or 2i+LIF conditions was extracted and reverse-transcribed with TaqMan primers specific to the indicated miRNAs, and then expression was profiled by qPCR. Error bars represent standard deviation from technical triplicate PCR reactions, and samples are normalized to a basket of reference small noncoding RNAs and the decrease in Ct values compared to v6.5 mESCs grown in serum+LIF is shown. See Methods for full details. **b**, Fold-change in expression of the indicated miRNAs in wild-type mESCs cultured in serum+LIF or 2i+LIF, induced and uninduced *let-7* mESCs cultured in serum+LIF, and MEFs grown under standard conditions. Changes are shown relative to v6.5 mESCs grown in serum+LIF. **c**, Comparison of genes that change in expression upon introduction of the ESCC miRNA miR-294 to *Dgcr8*^{-/-} mESCs in ref. 44 (*y* axis) to genes that change in expression between serum+LIF and 2i+LIF culture in single-cell RNA-seq data (*x* axis). Genes that are significantly upregulated in *Dgcr8*^{-/-} mESCs upon miR-294 introduction are indicated in blue, and those that are significantly downregulated are indicated in red. Selected genes upregulated by miR-294 and downregulated in 2i+LIF culture as compared to serum+LIF are

highlighted. As a group, genes downregulated by miR-294 show higher expression in 2i+LIF than in serum+LIF. **d**, mESC colonies staining uniformly positive for alkaline phosphatase show reduced levels of *Myc*. Overall and colony-specific *Myc* distribution in serum+LIF culture as measured by RNA FISH, showing uniformly positive (top), mixed (middle), or negative (bottom) alkaline phosphatase staining. **e**, Western blot showing reduction of Lin28a protein levels in iShLin28a cells upon addition of doxycycline. **f**, *let-7* expression changes in doxycycline-inducible *let-7* mESCs grown in serum+LIF upon induction. RT-qPCR was performed as in panel **a**, and Ct changes are shown relative to v6.5 mESCs in serum+LIF in a separate experiment from panel **a**. The inducible *let-7* construct is detected by the *let-7g* probe. **g**, Effect of miRNA transfection on self-renewal efficiency of *Dgcr8*^{-/-} mESCs. miRNA mimics were transfected into *Dgcr8*^{-/-} mESCs, and self-renewal efficiency was measured. Error bars indicate standard deviations between triplicate transfection experiments. Co-transfection of the ESCC miRNA miR-294 with a *let-7* miRNA results in enhanced self-renewal efficiency as compared to miR-294 alone. **h**, Expression changes of selected genes measured by qPCR upon culture of v6.5 mESCs in serum+LIF, 2i+LIF, or treatment with only Erk or GSK3 β inhibitors.

Extended Data Table 1 | List of genetic and chemical perturbations whose effects were profiled by single-cell qPCR

Perturbation	Effect	Cell line	Growth media	Feeders	Abbreviation	Ref.
Culture conditions						
Standard (v6.5 MEFs)	Baseline conditions	v6.5 mESCs	Serum+LIF	Yes	v6.5 on MEFs	
Feeder-free	Increased spontaneous differentiation	v6.5 mESCs	Serum+LIF	No	v6.5 FF	
Feeder-free (E14)	Increased spontaneous differentiation	E14Tg2A	Serum+LIF	No	E14 FF	
Ground state	Increased self-renewal efficiency	v6.5 mESCs	2i+LIF	No	V6.5 2i+LIF	13
Ground state (E14)	Increased self-renewal efficiency	E14Tg2A	2i+LIF	No	E14 2i+LIF	13
Genetic knockouts						
Polycomb repressive complex 2	Disruption of the PRC2 complex, loss of H3K27me3	<i>Eed</i> ^{-/-}	Serum+LIF	Yes	<i>Eed</i> ^{-/-}	31
DNA methyltransferases	Loss of DNA methylation	<i>Dnmt</i> TKO	Serum+LIF	Yes	<i>Dnmt</i> TKO	33
MicroRNA processing	Loss of mature miRNAs (Knockout of <i>Dgcr8</i>)	<i>Dgcr8</i> ^{-/-}	Serum+LIF	Yes	<i>Dgcr8</i> ^{-/-} on MEFs	34
				No	<i>Dgcr8</i> ^{-/-} FF	
	Loss of mature miRNAs (Knockout of <i>Dicer</i>)	<i>Dicer</i> ^{-/-}	Serum+LIF	Yes	<i>Dicer</i> ^{-/-} on MEFs	30
				No	<i>Dicer</i> ^{-/-} FF	
NuRD complex	Disruption of the NuRD complex	<i>Mbd3</i> ^{-/-}	Serum+LIF	Yes	<i>Mbd3</i> ^{-/-}	32
Chemical treatments						
Trichostatin A	HDAC inhibition	v6.5 mESCs	Serum+LIF	Yes	TSA	
Valproic acid	HDAC inhibition	v6.5 mESCs	Serum+LIF	Yes	VPA	
BIX01294	Inhibition of G9a histone methyltransferase	v6.5 mESCs	Serum+LIF	Yes	BIX	
Hoechst 33342	Cell cycle analysis (cells in G ₀ /G ₁ were isolated)	v6.5 mESCs	Serum+LIF	Yes	Hoechst	
Alternate cell types						
Induced pluripotent stem cells	Reprogrammed from somatic cells	iPSCs	Serum+LIF	Yes	iPSCs	
Epiblast stem cells	Primed state of pluripotency	EpiSCs	EpiSC growth media	Yes	EpiSCs	35

Structure of the *V. cholerae* Na⁺-pumping NADH:quinone oxidoreductase

Julia Steuber¹, Georg Vohl^{2,3*}, Marco S. Casutt^{2*}, Thomas Vorburger¹, Kay Diederichs⁴ & Günter Fritz²

NADH oxidation in the respiratory chain is coupled to ion translocation across the membrane to build up an electrochemical gradient. The sodium-translocating NADH:quinone oxidoreductase (Na⁺-NQR), a membrane protein complex widespread among pathogenic bacteria, consists of six subunits, NqrA, B, C, D, E and F. To our knowledge, no structural information on the Na⁺-NQR complex has been available until now. Here we present the crystal structure of the Na⁺-NQR complex at 3.5 Å resolution. The arrangement of cofactors both at the cytoplasmic and the periplasmic side of the complex, together with a hitherto unknown iron centre in the midst of the membrane-embedded part, reveals an electron transfer pathway from the NADH-oxidizing cytoplasmic NqrF subunit across the membrane to the periplasmic NqrC, and back to the quinone reduction site on NqrA located in the cytoplasm. A sodium channel was localized in subunit NqrB, which represents the largest membrane subunit of the Na⁺-NQR and is structurally related to urea and ammonia transporters. On the basis of the structure we propose a mechanism of redox-driven Na⁺ translocation where the change in redox state of the flavin mononucleotide cofactor in NqrB triggers the transport of Na⁺ through the observed channel.

Electrogenic NADH:ubiquinone oxidoreductases are central components of the respiratory chain in mitochondria and bacteria. These enzymes are integral membrane proteins that oxidize cytoplasmic NADH and reduce ubiquinone located in the membrane. The energy released by this redox reaction is used by complex I and the Na⁺-NQR to pump ions across the membrane, thus generating the electrochemical gradient that is essential for all living organisms¹. Although these two membrane protein complexes catalyse a very similar reaction, their basic architecture is completely different. The Na⁺-NQR is composed of the six subunits NqrA, B, C, D, E, F and has been reported to contain a unique set of cofactors: one FAD, a [2Fe-2S] cluster, two covalently bound flavin mononucleotide (FMN), riboflavin and a tightly bound ubiquinone^{2–4}. The Na⁺ gradient generated by the Na⁺-NQR is essential for *Vibrio cholerae* for being infectious, for example, by driving its monopolar flagellum required for motility. Moreover, the activity of Na⁺-NQR is linked to the expression of virulence factors, and inhibition of Na⁺-NQR decreases toxin production⁵. Genes for the Na⁺-NQR were identified in many pathogenic bacteria and have no eukaryotic homologue, designating the enzyme as a probable drug target in antibacterial therapy. Interestingly, four out of six subunits, NqrB, NqrC, NqrD and NqrE, are homologous to subunits of the so-called RNF (*Rhodobacter* nitrogen fixation) complex present in widely different groups of prokaryotes⁶ (Extended Data Figs 1, 2). The high similarity over four subunits suggests that the core machinery for Na⁺ translocation is identical in both respiratory complexes. To date no structural information is available for the Na⁺-NQR or the RNF complex. Here we report on the crystal structure of the entire Na⁺-NQR complex at 3.5 Å resolution (Extended Data Table 1). To gain further insights we complemented the low-resolution information with high-resolution structural data on several subunits, by determining the structures of the major soluble domains of subunits NqrA, NqrC and NqrF at 1.60 Å, 1.70 Å and 1.55 Å resolution, respectively.

Overall structure

The entire complex has dimensions of 87 Å × 138 Å × 52 Å (Fig. 1). The transmembrane part of Na⁺-NQR is mainly formed by the subunits

NqrB, NqrD and NqrE, which are integral membrane proteins with small domains in the cytoplasm or periplasm (Extended Data Fig. 3). NqrA, NqrC and NqrF are large hydrophilic subunits with the latter two anchored via a single transmembrane helix in the cytoplasmic membrane. The hydrophilic subunits NqrA and NqrF protrude into the cytoplasm, whereas NqrC resides in the periplasm (Fig. 1). NqrA, which lacks transmembrane helices, is tightly bound to NqrB, forming a complex with a large interaction surface (Fig. 1 and Extended Data Fig. 4a). Subunits NqrD and NqrE form a central structural unit that interacts with subunit NqrB on one side and the transmembrane helices of NqrC and NqrF on the opposed side (Extended Data Fig. 3d, e).

The hydrophilic subunits in the cytoplasm

NqrA consists of four domains (Extended Data Fig. 4b, c). The amino-terminal domain (residues 28–100) exhibits similarity to a biotin carboxyl carrier domain. It is followed by a Rossmann-fold domain (102–254), an ubiquitin-like domain (residues 258–329), and a carboxy-terminal helical domain (376–446). At the sequence level, NqrA shows no resemblance to proteins of known structure. However, a DALI search⁷ revealed an unexpected structural similarity with subunit Nqo1 of bacterial complex I⁸. Despite their distinct function, the similarity between NqrA and Nqo1 is remarkable with a root-mean-square deviation (r.m.s.d.) of 3.9 Å over 234 Cα positions (Extended Data Fig. 4d). The carboxy-terminal helical domain of NqrA shows similarity to a 2[4Fe-4S] cluster ferredoxin fold but does not contain an FeS centre. Although NqrA is structurally similar to subunit Nqo1 of complex I, there is no overlap at the functional level. Whereas Nqo1 binds an FMN cofactor in the Rossmann fold, NqrA contains no cofactor and it has been proposed that the subunit lost its cofactors during evolution of Na⁺-NQR⁹. Although we have recently mapped the ubiquinone interaction site in NqrA by photo-affinity labelling and NMR^{4,10}, we could not observe this cofactor in the low-resolution electron density of the entire Na⁺-NQR. NqrA comprises a deep solvent-accessible cavity that is formed by residues of the Rossmann-fold domain and the ubiquitin-like domain that is large enough to accommodate ubiquinone. NqrA forms major contacts with NqrB via its N and C

¹Department of Microbiology, Garbenstrasse 30, University of Hohenheim, 70599 Stuttgart, Germany. ²Institute for Neuropathology, University of Freiburg, Breisacher Strasse 64, 79106 Freiburg, Germany. ³Hermann-Staudinger-Graduate school, University of Freiburg, Hebelstrasse 27, 79104 Freiburg, Germany. ⁴Department of Biology, University of Konstanz, Universitätsstrasse 10, 78457 Konstanz, Germany.

*These authors contributed equally to this work.

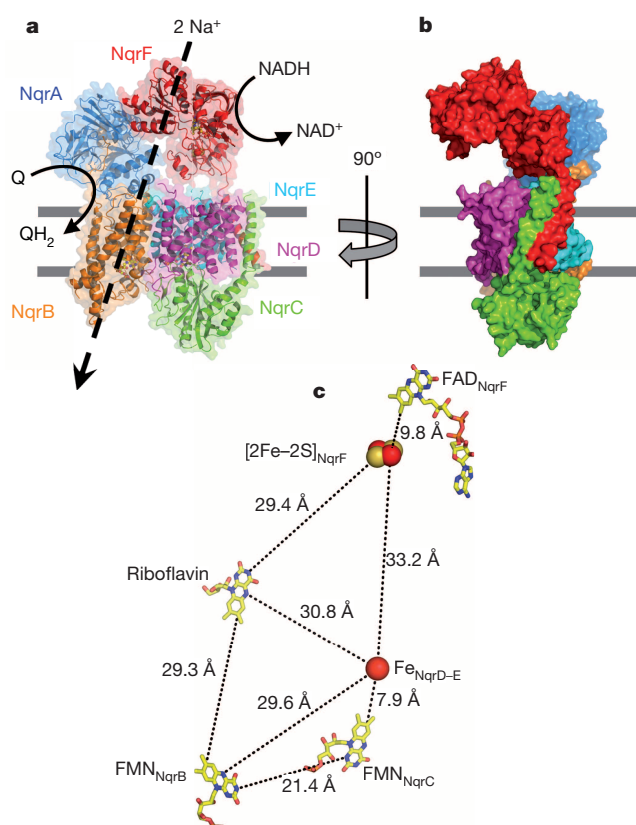


Figure 1 | Overall structure of Na^+ -NQR. **a, b,** Na^+ -NQR is composed of six subunits, NqrA–NqrF. While NqrA (blue) contains no transmembrane helix, NqrB (orange), NqrD (magenta) and NqrE (cyan) are integral membrane proteins. NqrF (red) and NqrC (green) are anchored in the cytoplasmic membrane by a single transmembrane helix. The membrane plane is indicated by grey lines. NADH oxidation occurs at NqrF and ubiquinone (Q) has been shown to bind to NqrA. The energy liberated is used to translocate two Na^+ per NADH oxidized³⁴. **b,** The solvent accessible surface of Na^+ -NQR is shown. **c,** The arrangement of the redox cofactors in Na^+ -NQR. Edge-to-edge distances of the cofactors are indicated by broken lines.

termini. The N-terminal residues 38–51 of NqrB reside in a deep groove of NqrA (Extended Data Fig. 4a). At the C terminus of NqrB, transmembrane helix 10 is elongated and protrudes into the cytoplasm, forming contacts with the C-terminal domain and the Rossmann-fold domain of NqrA.

NqrF consists of two major hydrophilic domains anchored by a single transmembrane helix to the Na^+ -NQR complex. The N-terminal ferredoxin-like domain following the transmembrane helix harbours a $[2\text{Fe}-2\text{S}]$ cluster and is located proximal to the membrane, whereas the C-terminal FAD-binding domain resides remote of the membrane (Fig. 1). The linker regions between the transmembrane helix, the ferredoxin-like domain and the FAD-binding domain were barely visible in the electron density, suggesting a high degree of flexibility. The FAD-binding domain of NqrF is reminiscent to the ferredoxin-NADP⁺ reductase-type flavoproteins and represents the NADH-oxidizing unit of Na^+ -NQR. In contrast to NqrA that exhibits multiple and large contacts to the integral membrane subunit NqrB, NqrF does not interact with the membrane subunits. Neither the FAD-binding nor the ferredoxin-like domain show any contact with the NqrB, NqrD or NqrE. The ferredoxin-like domain resides in a distance of 10 Å to NqrE and the FAD-binding domain in a distance of 25 Å to NqrD, leaving a large gap between the integral membrane subunits NqrD–E and NqrF. This spatial arrangement suggests that the hydrophilic domains of NqrF might move towards the membrane subunits of the complex, with the linker regions of NqrF acting as hinges.

NqrB harbours a Na^+ channel

NqrB comprises 10 transmembrane helices and the arrangement of the helices suggests that the subunit has originally evolved from gene duplication (Extended Data Fig. 3a). Search for structural homologues revealed that NqrB is homologous to urea and ammonium transporters from prokaryotes and eukaryotes. Despite the low sequence identity (5–12%), the structural similarity is rather high. In structural alignments of NqrB with a urea transporter from *Bos taurus* (PDB code 4EZC)¹¹, 205 residues were aligned with an r.m.s.d. of 3.7 Å for the α positions (Extended Data Fig. 5a). Correspondingly, structural alignment with ammonium transporters from different species revealed an r.m.s.d. of 3.6 Å for 210 residues^{12,13} (Extended Data Fig. 5b). Both urea transporter and ammonia transporter mediate the transport of small solutes^{11,14} across the membrane through a central cavity. The similarity of the three different membrane proteins is intriguing and suggests that they have evolved from a common ancestor that transported positively charged ions or molecules. In NqrB the central helices I, III, VI and VIII form a membrane-spanning channel (Fig. 2a, b). However, when compared to the urea transporter the channel in NqrB is narrowed by a relocation of the rather short helix VIII by approximately 5 Å towards the centre of the protein (Extended Data Fig. 5a, b). Several residues located on this helix appear to be crucial for the transport of Na^+ across the membrane. The channel opens at the cytoplasmic aspect of NqrB, forming a wide vestibule that displays a negatively charged surface (Extended Data Fig. 5c). Residue Asp 346 of NqrB is located at the bottom of this vestibule (Fig. 2c), and mutational studies have suggested that this aspartic residue is involved in Na^+ transport¹⁵. At a position halfway through the membrane the channel narrows sharply and the access to the second half of the channel, which opens to the periplasm, is blocked by side chains of hydrophobic residues: Phe 338 and Phe 342 located on helix VIII, Ile 164, Leu 165 and Leu 168 on helix III, and Val 64 of helix I (Fig. 2c, d). Rotation of the side chain of either Phe 342 or Ile 164 would completely block the observed channel. After this constriction the channel opens widely, forming a large cavity with an exit site to the periplasmic aspect of NqrB. In this arrangement Na^+ ions can enter the channel from the cytoplasmic or periplasmic side, but cannot pass the constriction. Interestingly, the cytoplasmic half-channel is negatively charged, whereas the periplasmic half-channel is positively charged (Extended Data Fig. 5c). This asymmetric charge distribution in the two half-channels might cause an enrichment of Na^+ ions in the cytoplasmic half versus the periplasmic half, facilitating the directional transport of the cations.

As a major difference to the urea and ammonium transporter proteins, NqrB comprises an additional periplasmic domain that harbours an FMN cofactor that is covalently linked via its phosphate group to Thr 236. This FMN-binding domain is assembled by insertions into the regions connecting the transmembrane helices I/II, V/VI and VII/VIII of NqrB and encompasses a total of 120 residues. The FMN itself is deeply buried in the protein (Fig. 2a) with the isoalloxazine moiety located within the transmembrane region and the cavity around the FMN connected to the periplasmic Na^+ half-channel (Extended Data Fig. 5c).

Thus, NqrB represents a molecule that preserved the basic architecture of a transporter but has acquired an additional and completely different function as a redox protein. An opening of the Na^+ channel must be coupled to changes in the redox state of the Na^+ -NQR complex and it appears reasonable that a change of the redox state of FMN in NqrB triggers Na^+ translocation as proposed previously on the basis of kinetic data¹⁶.

A Fe site in NqrD–E

The localization of the Na^+ channel in NqrB raises the question of the function of the other integral membrane subunits NqrD and NqrE. Structurally, both subunits are closely related and very likely the result of a gene duplication. Each subunit comprises six transmembrane helices connected by short loop regions in the cytoplasm and periplasm (Extended Data Figs 1a, b and 3b, c). As a major difference the transmembrane helices of both subunits exhibit an inverted path through the

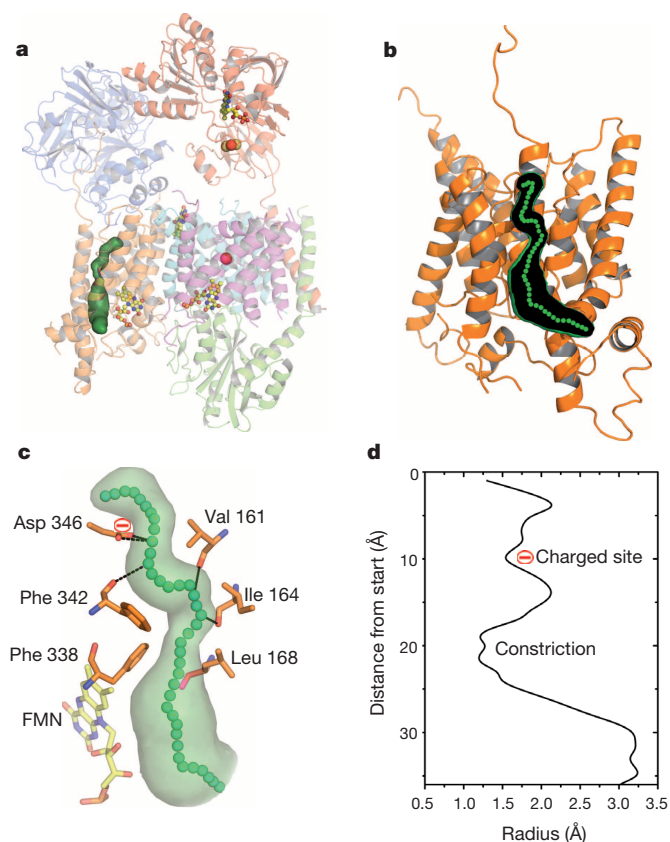


Figure 2 | Subunit NqrB harbours a Na⁺-channel. **a**, Analysis of the transmembrane subunits with CAVER revealed a hydrophilic channel (green) in subunit NqrB. **b**, Close-up view of the channel in NqrB. Transmembrane helix I is removed for clarity. The green spheres indicate the proposed path of the Na⁺ through NqrB. **c**, Key residues of the putative Na⁺ channel. The negatively charged side chain of Asp 346 can coordinate Na⁺ at the entry of the channel. The backbone carbonyl residues of Val 161, Ile 164 and Leu 168 located on transmembrane helix III can coordinate the Na⁺ on the predicted path. **d**, Ile 164, Leu 168 and on the opposite site Phe 338 and Phe 342 located on transmembrane helix VIII narrow the channel, forming a constriction.

membrane. NqrD and NqrE assemble into a symmetrical entity where both subunits are related by a rotation of 180° around a two-fold axis located in the centre of the membrane (Fig. 3b, c). Transmembrane helices I and IV from both subunits form the core of the NqrD–E unit. These four transmembrane helices are organized as half-helices with each half-helix tilted by approximately 45° with respect to the membrane plane. The linker regions of the four central helices meet in the centre, thereby forming a core structure resembling an hourglass. In the centre of this core unit a strong anomalous signal was observed originating from a Fe centre (Fig. 3a, d). The identity of Fe was confirmed by comparison of the intensity of the peak at different wavelengths (Extended Data Table 2). Four cysteine residues originating from each of the core helices coordinate the Fe via the thiol group of the side chain. It appears that the four transmembrane helices harbouring the cysteine residues unwind in the centre of the membrane to accommodate for iron coordination (Fig. 3b, d). The coordination and localization of the single Fe in the membrane region of Na⁺-NQR is unprecedented and represents a so far unknown redox cofactor of Na⁺-NQR.

NqrC represents a new type of flavoprotein

NqrC contains, like NqrB, a covalently bound FMN that is linked via a phosphoester to the side chain of Thr 225 (ref. 17). There is a debate on the topology of NqrC. On the basis of PhoA fusions it was proposed that NqrC resides in the cytoplasm¹⁸, whereas the insertion of the covalently bound FMN into NqrC in the periplasm suggested a periplasmic

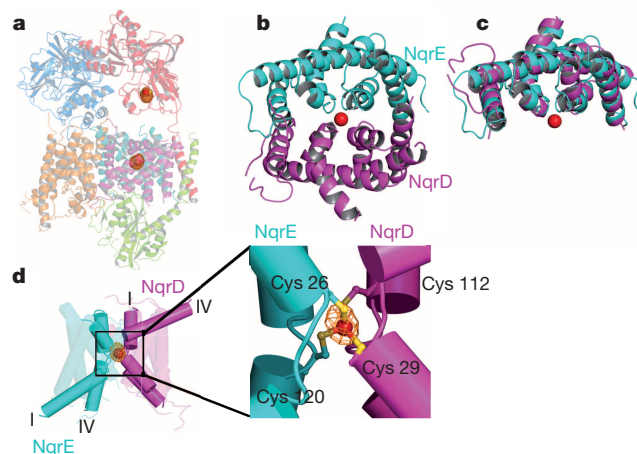


Figure 3 | Subunits NqrD and NqrE contain a novel Fe centre. **a**, Two strong peaks in the anomalous difference map shown in orange at 5σ were detected. The first peak was assigned to the [2Fe-2S] centre in NqrF and the second peak in the centre between subunit NqrD and NqrE. The [2Fe-2S] cluster and the Fe are shown as spheres. **b**, **c**, Top view of subunits NqrD (magenta) and NqrE (cyan). NqrD and NqrE form a symmetrical dimer. Both subunits are related by a two-fold axis. The centre of this unusual dimer is formed by transmembrane helices I and IV of NqrD and NqrE, which form two half-helices instead of a continuous transmembrane helix. **d**, Approximately in the centre of the membrane plane the helices unfold and four Cys residues originating from each helix coordinate the Fe centre. The anomalous difference map is shown at 8σ.

localization¹⁹. Here we show that NqrC resides in the periplasm (Figs 1 and 4). The single transmembrane helix that anchors this hydrophilic subunit in the membrane elongates to a long helix that is part of the fold of NqrC. Structurally, NqrC is not related to any known protein family. The fold of NqrC comprises a central five-strand β-sheet that is sandwiched on both sides by two helices from the N terminus and C terminus. At the periplasmic aspect, NqrC plunges into a deep cavity of the NqrD–E unit (Fig. 4a, b). The FMN of NqrC is covalently attached via a phosphoester bond to the oxygen of the side chain of Thr 225 located in a helix of the C terminus (Fig. 4b). The isoalloxazine moiety of the FMN resides in an open pocket and more than half of the isoalloxazine plane protrudes out from the NqrC protein matrix to be embedded in the cavity formed by NqrD and NqrE (Fig. 4b, c).

Riboflavin binding site

Deletion of subunit NqrB leads to the loss of riboflavin in Na⁺-NQR, indicating that riboflavin is bound to this subunit²⁰. After completion of the model of Na⁺-NQR we did not observe any additional difference electron density inside NqrB that could encompass a riboflavin cofactor. An unbiased analysis of the $F_o - F_c$ difference map around the entire Na⁺-NQR by automatic fitting²¹ revealed a large patch of $F_o - F_c$ density at the interface between NqrB and NqrE. The aromatic isoalloxazine moiety fits well into the $F_o - F_c$ density (Extended Data Fig. 6), whereas other candidates, like for example ubiquinone, are too small. The riboflavin-binding site is located in the transmembrane region but close to the cytoplasmic aspect, suitable for electron transfer to ubiquinone located in NqrA^{4,10}.

Electron transfer pathway

In total there are seven redox cofactors described to be present in Na⁺-NQR: four flavin cofactors, one [2Fe-2S] cluster, the Fe site in NqrD–E described here for the first time (Fig. 1c), and a tightly bound ubiquinone. Edge-to-edge and centre-to-centre distances of the redox cofactors are listed in Extended Data Table 3. NADH is oxidized in a two-electron step at the FAD of subunit NqrF and electrons are transferred one-by-one from FAD to the [2Fe-2S] cluster residing at an edge-to-edge distance of 9.8 Å in the N-terminal domain of NqrF. Mutational and fast

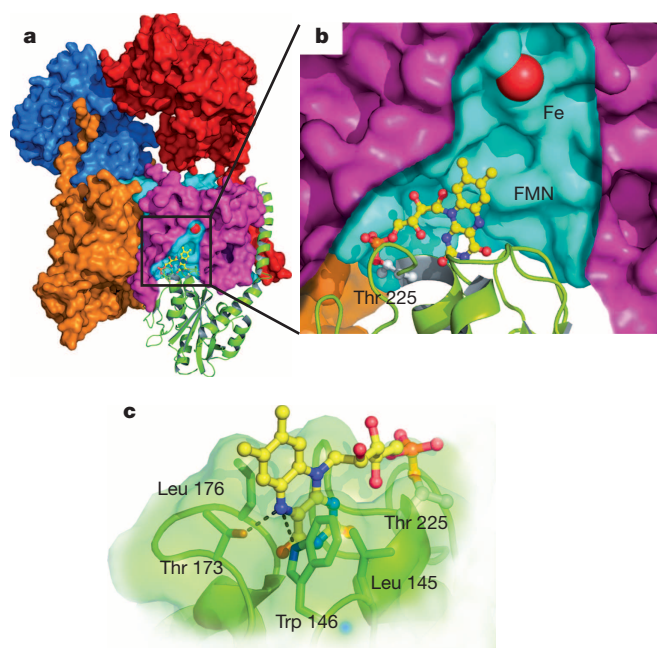
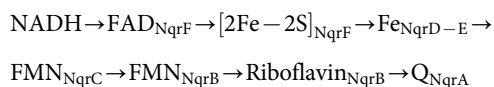


Figure 4 | Subunit NqrC exhibits a new flavoprotein fold and FMN-binding motif. **a**, A part of subunit NqrD (magenta) is removed to show the localization of NqrC (green). **b**, The FMN_{NqrC} is covalently bound via a phosphoester bond between the ribityl phosphate and the Thr 225 side chain. The distance between the FMN and the Fe centre in NqrD–E is just 8 Å, with the C7 and C8 methyl groups of the isoalloxazine pointing directly to the Fe centre. **c**, NqrC binds FMN in an unprecedented way; only the pyrimidine moiety of isoalloxazine is embedded in the protein matrix with the benzene moiety protruding from the surface. The aromatic isoalloxazine is sandwiched between Leu 145 and Leu 176. Additional hydrogen bonds by the side chains of Trp 146 and Thr 173 to N5 of FMN might stabilize its position and tune the redox potential. A further hydrogen bond between O4 of the ribityl and N1 might stabilize the radical state of the FMN.

kinetics data indicate that the electrons are subsequently transferred to FMN_{NqrC}¹⁶, and via FMN_{NqrB}, and riboflavin_{NqrB} finally to ubiquinone. However, FMN_{NqrC} resides in the periplasm and the electrons have to cross the cytoplasmic membrane, making direct electron transfer from NqrF to NqrC very unlikely. The Fe site in the core of subunits NqrD and NqrE is well suited to bridge electron transfer between NqrF and NqrC. The proposed electron transfer pathway can be summarized in the following scheme:



However, several distances between the redox centres in the proposed electron transfer pathway are too large for fast electron transfer²² and are not in agreement with measured rates of Na⁺-NQR^{16,23}. The edge-to-edge distance between the [2Fe-2S] cluster in NqrF and the Fe in NqrD–E is 33.4 Å, whereas the isoalloxazine moiety of FMN_{NqrC} is located in very close proximity to the Fe_{NqrD-E} with an edge-to-edge distance of just 7.9 Å. During turnover the N-terminal domain of NqrF has to undergo a large conformational change decreasing the gap between [2Fe-2S]_{NqrF} and Fe_{NqrD-E} to allow for efficient electron transfer. Such a conformational change has been observed so far only in *bc*₁ complex of the respiratory chain. The [2Fe-2S] cluster bearing subunit of *bc*₁ complex switches between two positions to shuttle electrons from ubiquinol to cytochrome *c*₁²⁴. In subunit NqrF the transmembrane helix, the [2Fe-2S] domain and the FAD-binding domain are connected by flexible linkers, enabling a rotational movement of the [2Fe-2S] domain towards the membrane plane and shortening the distance between [2Fe-2S]_{NqrF} and Fe_{NqrD-E}.

From FMN_{NqrC} electron transfer proceeds to the covalent FMN in NqrB. FMN_{NqrB} is bound to Thr 236 and, like FMN_{NqrC}, it is located at the periplasmic side of the complex, close to the hydrophobic core of NqrB. The edge-to-edge distance between FMN_{NqrC} and FMN_{NqrB} is 21.4 Å, but might decrease to approximately 15 Å according to electron paramagnetic resonance (EPR) analysis of spin–spin interactions between the radical state of both FMN cofactors²⁵. Both, FMN_{NqrC} and FMN_{NqrB} are well shielded from the solvent in order to minimize side reactions in the periplasm. Riboflavin was localized at an edge-to-edge distance of 29.3 Å to FMN_{NqrB} in the hydrophobic/hydrophilic transition zone at the cytoplasmic side of NqrB. Again this distance is too large for fast electron transfer. The observed electron transfer from FMN_{NqrB} to riboflavin_{NqrB} requires either a movement of FMN or riboflavin in NqrB. The covalent attachment of FMN to Thr 236 makes a movement unlikely, suggesting the possibility for a switch involving Riboflavin_{NqrB}.

Taking into account the data from previous studies and the now available structural data, two passages of the electron across the cytoplasmic membrane must be predicted. Such a pathway has implications on the mechanism and energetics of Na⁺ translocation as discussed later. Moreover, the architecture of the Na⁺-NQR reveals that NqrB, NqrC, NqrD and NqrE comprise the core machinery of the pump and carry the redox cofactors required for redox-driven Na⁺ translocation. Subunits NqrA and NqrF represent electron input and electron output devices and are substituted in the RNF complex by functionally different subunits. Noteworthy, NqrB, NqrC, NqrD and NqrE are highly homologous to subunits RnfD, RnfG, RnfA and RnfE of the RNF complex^{6,26} (Extended Data Figs 1 and 2), underlining their central function, and strongly suggesting that changes in the redox state of cofactors within these subunits must be critically coupled to Na⁺ translocation events.

Mechanism of Na⁺ translocation

On the basis of the available data we developed a model of coupling between electron transfer and Na⁺ translocation in Na⁺-NQR. Two general but different mechanisms describing the coupling of electron transfer to Na⁺ transport in Na⁺-NQR have been proposed previously. In the first model a mechanism of thermodynamic coupling where Na⁺ binds to a cofactor dependent on its redox state was suggested²⁷. In such a scenario co-translocation of an electron and Na⁺ across the membrane would be electroneutral and therefore energetically much less unfavourable. This model predicts that the redox potentials of cofactors involved are dependent on the presence of the coupling ion. However, so far there was no clear hint for Na⁺-dependent changes in redox potential^{28,29}. In a second model no direct interaction of Na⁺ with a redox cofactor occurs and Na⁺ translocation is mediated by redox-dependent conformational changes. The data of previous studies and the data provided here support the latter model and link distinct electron transfer steps to ion translocation events.

The oxidation of NADH at FAD_{NqrF} and the subsequent electron transfer to [2Fe-2S]_{NqrF} (Fig. 5a) are presumably not coupled to ion translocation events, as supported by kinetic data³⁰ showing that both reactions are independent on the presence of the coupling ion Na⁺. The transition of an electron from [2Fe-2S]_{NqrF} to Fe_{NqrD-E} triggers most likely binding of a Na⁺ in the negatively charged binding pocket in NqrB that opens to the cytoplasm (Fig. 5a). This is in agreement with kinetic data showing that electron transfer from [2Fe-2S]_{NqrF} to FMN_{NqrC} is dependent on Na⁺ and triggers capture of Na⁺^{16,31}. Both, the Fe_{NqrD-E} and the Na⁺ binding site at Asp 346 in the pocket of NqrB are about half-way through the membrane, representing an electroneutral and energetically favoured relay station of the transport. The conformational change in NqrF required for electron transfer between [2Fe-2S]_{NqrF} to Fe_{NqrD-E} might be coupled to subtle changes in NqrB facilitating access and binding of Na⁺. The short distance between Fe_{NqrD-E} and FMN_{NqrC} promotes a very fast electron transfer in the subsequent step that is presumably not coupled to the comparable slow ion translocation events. It is reasonable to assume that the next electron transfer step from FMN_{NqrC} to FMN_{NqrB} triggers the occlusion of the Na⁺ in NqrB (Fig. 5a). In

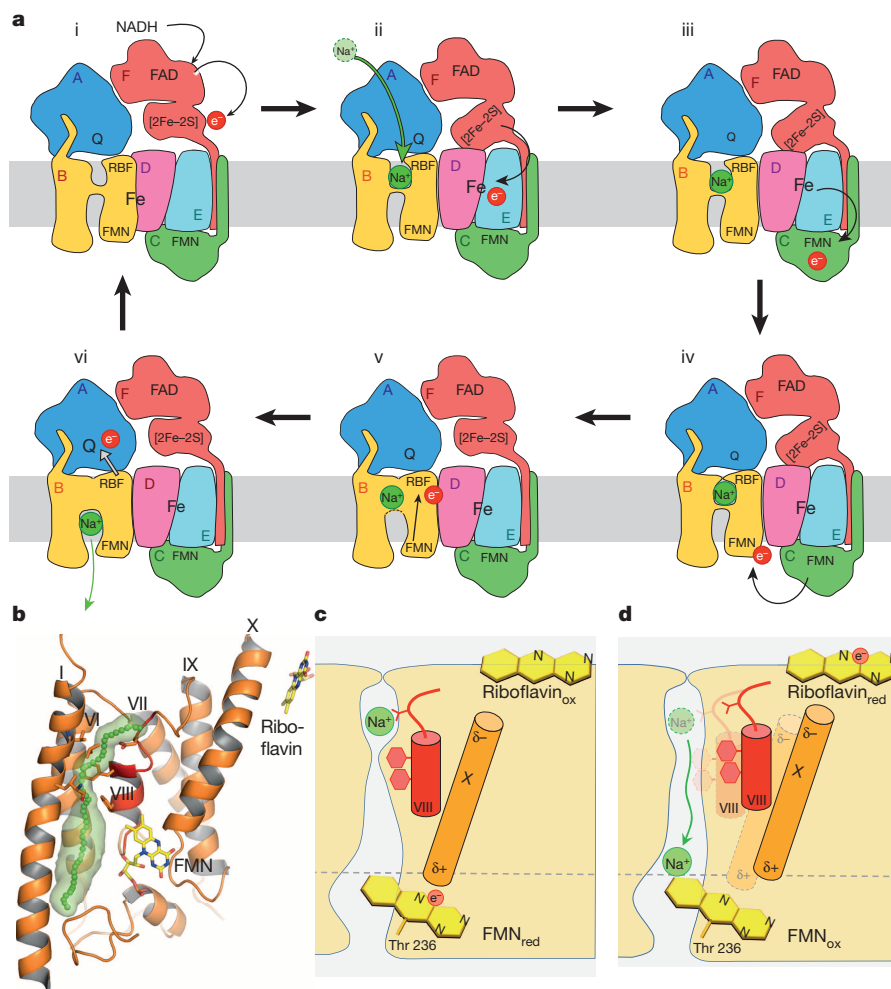


Figure 5 | Mechanism of redox-driven Na⁺ translocation. **a**, Model for coupling between electron transfer and Na⁺ translocation. NADH is oxidized at FAD_{NqrF} that mediates single electron transfer to [2Fe-2S]_{NqrF} (i). A conformational change in NqrF is required to transfer the electron to the Fe_{NqrD-E}. This step might facilitate concomitant Na⁺-binding in the cytoplasmic half-channel of NqrB (ii). Subsequently the electron is transferred to FMN_{NqrC} (iii), and FMN_{NqrB} (iv). The latter step presumably triggers the occlusion of Na⁺. Electron transfer from FMN_{NqrB} to riboflavin_{NqrB} (RBF; v) leads to an opening of the constriction and Na⁺ is released to the periplasmic

agreement with this assumption, EPR-based distance measurements between the one-electron-reduced FMN cofactors point towards substantial Na⁺-dependent conformational changes²⁵. The energy of this step might be further used to strip off the water shell of the Na⁺ and coordinate the ion by oxygen atoms deriving from the protein backbone and from the NqrB Asp 346 side chain. The redox state of flavin cofactors is known to control protein function by acting as a redox switch inducing conformational changes^{32,33}. In NqrB the one-electron reduced state of FMN_{NqrB} might pull transmembrane helix X pointing with its helix dipole directly towards the N5 of FMN_{NqrB} (Fig. 5b, c). Transmembrane helix X of NqrB is in close contact with the short transmembrane helix VIII carrying residues Phe342 and Phe338, which constricts the proposed Na⁺ channel (Figs 2c and 5b, c). Any movement of helix X would induce a movement of helix VIII as well. Previous studies¹⁶ put forward that electron transfer from FMN_{NqrB} to riboflavin_{NqrB} is coupled to the release of Na⁺ to the periplasm. Oxidation of the FMN_{NqrB} radical to the uncharged form might lead to the loss of the interaction between N5 of FMN_{NqrB} and the helix dipole, thereby releasing helix X. Concomitantly, helix VIII moves along with helix X. This movement opens the constriction involving residues Phe 338 and Phe 342 and triggers translocation of the Na⁺ to the periplasmic half-channel (Fig. 5d).

half channel of NqrB. The release might be dependent on the reduction of ubiquinone (vi). **b–d**, The redox state of FMN_{NqrB} controls the position of helix X pointing with the positive end of the helix dipole directly towards the N5. Electron transfer from FMN_{NqrB} to riboflavin_{NqrB} releases helix X that will move along with helix VIII. The shift of Phe 338 and Phe 342 located on helix VIII of NqrB will open the constriction and trigger translocation of the Na⁺ to the periplasmic half-channel. The wide periplasmic half-channel is most likely filled with water molecules, which will immediately form a water shell around the Na⁺.

The wide and solvent-accessible periplasmic half-channel is most likely filled with water molecules, which will immediately form a hydration shell around the Na⁺, making this step energetically favourable. According to this model the subsequent final electron transfer from riboflavin_{NqrB} to ubiquinone is not essential for Na⁺ translocation.

The structure of Na⁺-NQR of human pathogen *Vibrio cholerae* presented here suggests a new mechanism for the coupling between redox reactions and ion translocation events in a respiratory complex. Na⁺-NQR is a central component of the energy conservation system in a plethora of pathogenic bacteria. This structure may serve as the basis for the development of new antibiotics against the worldwide increase in antibiotic-resistant pathogens.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 August; accepted 24 October 2014.

- Mitchell, P. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **191**, 144–148 (1961).
- Pfenniger-Li, X. D., Albracht, S. P., van Belzen, R. & Dimroth, P. NADH:ubiquinone oxidoreductase of *Vibrio alginolyticus*: purification, properties, and reconstitution of the Na⁺ pump. *Biochemistry* **35**, 6233–6242 (1996).

3. Juárez, O., Nilges, M. J., Gillespie, P., Cotton, J. & Barquera, B. Riboflavin is an active redox cofactor in the Na⁺-pumping NADH:quinone oxidoreductase (Na⁺-NQR) from *Vibrio cholerae*. *J. Biol. Chem.* **283**, 33162–33167 (2008).
4. Casutt, M. S. *et al.* Localization of ubiquinone-8 in the Na⁺-pumping NADH:quinone oxidoreductase from *Vibrio cholerae*. *J. Biol. Chem.* **286**, 40075–40082 (2011).
5. Häse, C. C. & Mekalanos, J. J. Effects of changes in membrane sodium flux on virulence gene expression in *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **96**, 3183–3187 (1999).
6. Müller, V., Imkamp, F., Biegel, E., Schmidt, S. & Dilling, S. Discovery of a ferredoxin:NAD⁺-oxidoreductase (Rnf) in *Acetobacterium woodii*. *Ann. NY Acad. Sci.* **1125**, 137–146 (2008).
7. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
8. Baradaran, R., Berrisford, J. M., Minhas, G. S. & Sazanov, L. A. Crystal structure of the entire respiratory complex I. *Nature* **494**, 443–448 (2013).
9. Reyes-Prieto, A., Barquera, B. & Juárez, O. Origin and evolution of the sodium-pumping NADH: ubiquinone oxidoreductase. *PLoS ONE* **9**, e96696 (2014).
10. Nedielkov, R., Steffen, W., Steuber, J. & Möller, H. M. NMR reveals double occupancy of quinone-type ligands in the catalytic quinone binding site of the Na⁺-translocating NADH:quinone oxidoreductase from *Vibrio cholerae*. *J. Biol. Chem.* **288**, 30597–30606 (2013).
11. Levin, E. J. *et al.* Structure and permeation mechanism of a mammalian urea transporter. *Proc. Natl Acad. Sci. USA* **109**, 11194–11199 (2012).
12. Khademi, S. *et al.* Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 Å. *Science* **305**, 1587–1594 (2004).
13. Andrade, S. L., Dickmanns, A., Ficner, R. & Einsle, O. Crystal structure of the archaeal ammonium transporter Amt-1 from *Archaeoglobus fulgidus*. *Proc. Natl Acad. Sci. USA* **102**, 14994–14999 (2005).
14. Wacker, T., Garcia-Celma, J. J., Lewé, P. & Andrade, S. L. Direct observation of electrogenic NH₄⁺ transport in ammonium transport (Amt) proteins. *Proc. Natl Acad. Sci. USA* **111**, 9995–10000 (2014).
15. Juárez, O., Athearn, K., Gillespie, P. & Barquera, B. Acid residues in the transmembrane helices of the Na⁺-pumping NADH:quinone oxidoreductase from *Vibrio cholerae* involved in sodium translocation. *Biochemistry* **48**, 9516–9524 (2009).
16. Juárez, O., Morgan, J. E., Nilges, M. J. & Barquera, B. Energy transducing redox steps of the Na⁺-pumping NADH:quinone oxidoreductase from *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **107**, 12505–12510 (2010).
17. Casutt, M. S., Schlosser, A., Buckel, W. & Steuber, J. The single NqrB and NqrC subunits in the Na⁺-translocating NADH:quinone oxidoreductase (Na⁺-NQR) from *Vibrio cholerae* each carry one covalently attached FMN. *Biochim. Biophys. Acta* **1817**, 1817–1822 (2012).
18. Duffy, E. B. & Barquera, B. Membrane topology mapping of the Na⁺-pumping NADH:quinone oxidoreductase from *Vibrio cholerae* by PhoA/GFP fusion analysis. *J. Bacteriol.* **188**, 8343–8351 (2006).
19. Bertsova, Y. V. *et al.* Alternative pyrimidine biosynthesis protein ApbE is a flavin transferase catalyzing covalent attachment of FMN to a threonine residue in bacterial flavoproteins. *J. Biol. Chem.* **288**, 14276–14286 (2013).
20. Casutt, M. S. *et al.* Localization and function of the membrane-bound riboflavin in the Na⁺-translocating NADH:quinone oxidoreductase (Na⁺-NQR) from *Vibrio cholerae*. *J. Biol. Chem.* **285**, 27088–27099 (2010).
21. Terwilliger, T. C., Klei, H., Adams, P. D., Moriarty, N. W. & Cohn, J. D. Automated ligand fitting by core-fragment fitting and extension into density. *Acta Crystallogr. D* **62**, 915–922 (2006).
22. Page, C. C., Moser, C. C., Chen, X. & Dutton, P. L. Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Nature* **402**, 47–52 (1999).
23. Tao, M., Casutt, M. S., Fritz, G. & Steuber, J. Oxidant-induced formation of a neutral flavosemiquinone in the Na⁺-translocating NADH:quinone oxidoreductase (Na⁺-NQR) from *Vibrio cholerae*. *Biochim. Biophys. Acta* **1777**, 696–702 (2008).
24. Zhang, Z. *et al.* Electron transfer by domain movement in cytochrome bc₁. *Nature* **392**, 677–684 (1998).
25. Verkhovsky, M. I. *et al.* Sodium-dependent movement of covalently bound FMN residue(s) in Na⁺-translocating NADH:quinone oxidoreductase. *Biochemistry* **51**, 5414–5421 (2012).
26. Steuber, J. *et al.* Central role of the Na⁺-translocating NADH:quinone oxidoreductase (Na⁺-NQR) in sodium bioenergetics of *Vibrio cholerae*. *Biol. Chem.* **395**, 1389–1399 (2014).
27. Rich, P. R., Meunier, B. & Ward, F. B. Predicted structure and possible ionmotive mechanism of the sodium-linked NADH-ubiquinone oxidoreductase of *Vibrio alginolyticus*. *FEBS Lett.* **375**, 5–10 (1995).
28. Bogachev, A. V., Bertsova, Y. V., Aitio, O., Permi, P. & Verkhovsky, M. I. Redox-dependent sodium binding by the Na⁺-translocating NADH:quinone oxidoreductase from *Vibrio harveyi*. *Biochemistry* **46**, 10186–10191 (2007).
29. Verkhovsky, M. I. & Bogachev, A. V. Sodium-translocating NADH:quinone oxidoreductase as a redox-driven ion pump. *Biochim. Biophys. Acta* **1797**, 738–746 (2010).
30. Bogachev, A. V., Belevich, N. P., Bertsova, Y. V. & Verkhovsky, M. I. Primary steps of the Na⁺-translocating NADH:ubiquinone oxidoreductase catalytic cycle resolved by the ultrafast freeze-quench approach. *J. Biol. Chem.* **284**, 5533–5538 (2009).
31. Bogachev, A. V., Bertsova, Y. V., Ruuge, E. K., Wikstrom, M. & Verkhovsky, M. I. Kinetics of the spectral changes during reduction of the Na⁺-motive NADH:quinone oxidoreductase from *Vibrio harveyi*. *Biochim. Biophys. Acta* **1556**, 113–120 (2002).
32. Zhu, W. & Becker, D. F. Flavin redox state triggers conformational changes in the PutA protein from *Escherichia coli*. *Biochemistry* **42**, 5469–5477 (2003).
33. Leung, K. K. & Shilton, B. H. Chloroquine binding reveals flavin redox switch function of quinone reductase 2. *J. Biol. Chem.* **288**, 11242–11251 (2013).
34. Bogachev, A. V., Murtazina, R. A. & Skulachev, V. P. The Na⁺/e⁻ stoichiometry of the Na⁺-motive NADH:quinone oxidoreductase in *Vibrio alginolyticus*. *FEBS Lett.* **409**, 475–477 (1997).

Acknowledgements We thank the staff at beamlines X06SA and X06DA at Swiss Light Source for excellent support. This work was supported by contract research 'Methoden in den Lebenswissenschaften' of the Baden-Württemberg Stiftung P-LS-Meth/4 (to J.S. and G.F.), and by the Deutsche Forschungsgemeinschaft grant FR 1321/3-1 (to J.S.) and grant FR 1488/3-2 (to G.F.). We thank Y. Obermaier for expression and preparation of Na⁺-NQR.

Author Contributions J.S., G.F. and T.V. developed expression constructs; J.S. and M.S.C. developed purification procedures; T.V., M.S.C. and G.V. expressed the protein; M.S.C., G.V. and G.F. purified the entire complex or single subunits; M.S.C., G.V. and G.F. performed crystallization, crystal harvesting and data collection. G.F. and K.D. performed data processing and determination of phases. G.F. performed model building and refinement. G.F. prepared the figures. G.F. and J.S. wrote the manuscript.

Author Information Coordinates and structure factors for the entire complex of Na⁺-NQR and of individual subunits NqrA_{1–377}, NqrC_{33–257} and NqrF_{129–408} have been deposited in the Protein Data Bank. The PDB accession codes are 4P6V (entire NQR complex), 4U9O (subunit NqrA, crystal 1), 4U9Q (subunit NqrA, crystal 2), 4U9S (subunit NqrC), and 4U9U (subunit NqrF). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.F. (guenter.fritz@uniklinik-freiburg.de).

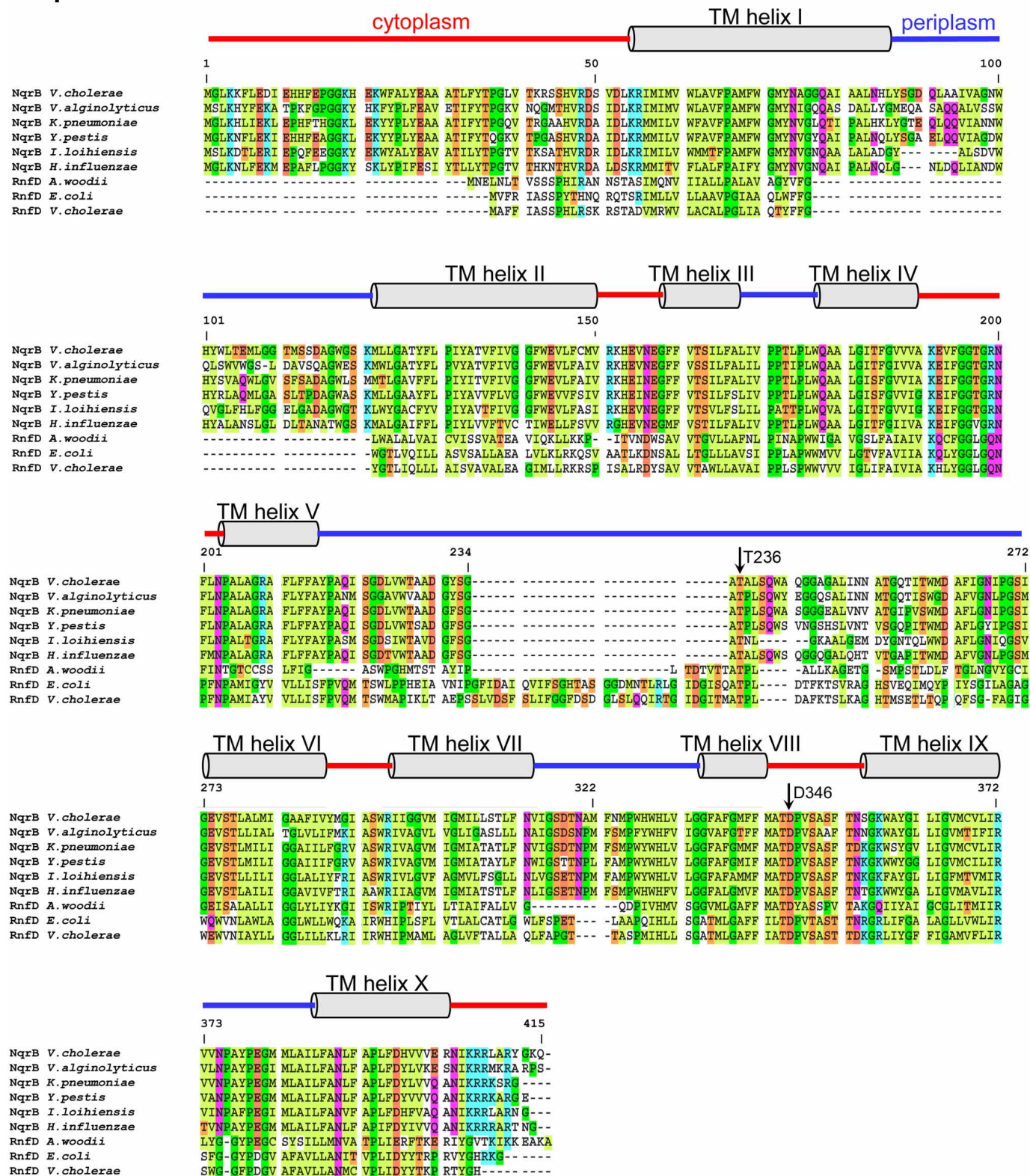
METHODS

Na⁺-NQR was expressed, purified and crystallized as described previously³⁵. Selenomethionine (SeMet)-labelled Na⁺-NQR was obtained by growth in minimal medium with L-selenomethionine as sole source for methionine as described in the following. From glycerol stocks, *V. cholerae* O395-N1 Δnqr³⁶, transformed with pNQR1²³ or pNqr-PreScission (this work), was cultivated on LB plates containing 10 g l⁻¹ tryptone, 5 g l⁻¹ yeast extract, 10 g l⁻¹ NaCl, 15 g l⁻¹ bactoagar, 50 mM potassium phosphate (KP_i) pH 8.5, 10 mM glucose, 50 µg ml⁻¹ streptomycin and 200 µg ml⁻¹ ampicillin. A single colony was used to inoculate 5 ml LB (composition as described above, but without bactoagar), and cells were grown for 16 h at 37 °C, 220 r.p.m. shaking. Growth was continued in 150 ml LB (inoculum size: 3 ml) for 4 h at 37 °C, 220 r.p.m. shaking. 15 ml were used to inoculate 1 litre Se-Met minimal medium (in a 5 l Erlenmeyer flask) consisting of 15 mM (NH₄)₂SO₄, 50 mM KP_i pH 8.5, 0.8 µM FeSO₄, alanine, arginine, glycine, glutamine, histidine, isoleucine, leucine, lysine, proline, serine, threonine, valine, asparagine, aspartic acid, cysteine, glutamic acid, phenylalanine, tyrosine, hypoxanthine, uracil (40 µg ml⁻¹ of each), tryptophan (20 µg ml⁻¹), *p*-hydroxybenzoic acid (4 µg ml⁻¹), biotin, nicotinamide, thiamine (2 µg ml⁻¹ of each), riboflavin (1 µg ml⁻¹), 1 mM MgSO₄, 10 mM glucose, 1% glycerol (v/v), 1% DMSO (v/v), 170 mM NaCl, 50 µg ml⁻¹ streptomycin, 200 µg ml⁻¹ ampicillin and L-selenomethionine (50 µg ml⁻¹). At 37 °C, 150 r.p.m. shaking, cells were grown to an optical density at 600 nm of 1.0. Na⁺-NQR production was induced by adding 5 ml 20% L-arabinose, and expression was continued for 12 h at 30 °C, 110 r.p.m. shaking. The culture was cooled on ice, cells were collected by centrifugation (10,000g, 30 min, 4 °C), washed in 100 ml lysis buffer (500 mM NaCl, 50 mM sodium phosphate pH 8.0), suspended in lysis buffer (1 g wet weight of cells per ml of buffer), frozen in liquid nitrogen and stored at -80 °C. After an expression phase of 12 h, cultures in selenomethionine minimal medium typically reached an optical density at 600 nm of 1.8. Typical yields were 4.5 g cells wet weight per litre of culture. The (SeMet)-labelled Na⁺-NQR was purified as described³⁵. Diffraction data were collected at the Swiss Light Source, beamline X06SA equipped with a PILATUS 6 M detector. All diffraction data were integrated using XDS³⁷. A Ta₆Br₁₂ derivative was obtained by soaking native crystals overnight in 2 mM Ta₆Br₁₂ (Extended Data Table 1). Initial phases were calculated with SHARP³⁸. After solvent flattening with Solomon³⁹ or DM⁴⁰ the resulting electron density map was of sufficient quality to build about 18 transmembrane helices. However, it was not possible to place any further structural elements into the experimental map. Further phase information was obtained from a SeMet derivative (Extended Data Table 1). The initial transmembrane helix model was used as a search model in a combined molecular replacement SAD approach using Phaser⁴¹. A total of 58 out of 70 Se sites were found in the first trial and the resulting electron density map was of good quality. Several cycles of model building with COOT⁴² and SAD phasing were performed. A large part of the sequence of the transmembrane helices could be assigned using the positions of the Se sites in methionine residues as marker. The previously proposed symmetric and inverted topology of NqrD and NqrE helped to assign the helices to the different subunits. However, several transmembrane helices of NqrB contain no methionine and it was not possible to trace the sequence throughout the ten transmembrane helices. We therefore performed sequence-based searches for structures homologous to the different subunits of Na⁺-NQR using the HHPred server⁴³ that utilizes hidden Markov models. NqrB was predicted to be structurally homologous to the urea transporter¹¹ and the structure could be aligned with the transmembrane helices guiding the tracing of the sequence. The low resolution of the experimental map allowed the unambiguous placement of side chains for methionine or tryptophan residues only. We therefore produced recombinantly several soluble domains of NqrA, NqrC and NqrF, crystallized these proteins separately and solved their structures at 1.60 Å, 1.70 Å and 1.55 Å^{44,45}. The structures of these subunits were placed into the low resolution experimental map by molecular replacement using MOLREP⁴⁶, and their localization in the complex was verified by the positions of the methionine residues indicated by the Se positions. Refinement was performed using Refmac5⁴⁷ and with phenix.refine⁴⁸ using the high-resolution structures of NqrA, NqrC and NqrF as reference models. The

high-resolution limits were established by paired-refinement⁴⁹. Several cycles of manual rebuilding and refinement allowed finally the placement of 1,774 residues and 6 cofactors and the structure was refined at 3.5 Å resolution to a final *R*_{work} = 24.6% and *R*_{free} = 27.6% (Extended Data Table 1) with good stereochemistry parameters compared to other structures determined at this resolution. The refined model at 3.5 Å resolution showed no substantial deviation from the high resolution structural data of the single subunits (Extended Data Table 4). Analysis with MolProbity⁵⁰ shows that 90.2% of the residues in Na⁺-NQR are in the favourable regions of the Ramachandran plot with 7% in the allowed regions, and 2.8% outliers. Most of the outliers are located in the integral membrane subunits or areas where the electron density was less well defined. Cavities and putative channels were calculated with CAVER⁵¹, and surface potential was calculated with APBS⁵². All figures were prepared with PyMol⁵³.

35. Casutt, M. S., Wendelspiess, S., Steuber, J. & Fritz, G. Crystallization of the Na⁺-translocating NADH:quinone oxidoreductase from *Vibrio cholerae*. *Acta Crystallogr. F* **66**, 1677–1679 (2010).
36. Barquera, B. *et al.* Purification and characterization of the recombinant Na⁺-translocating NADH:quinone oxidoreductase from *Vibrio cholerae*. *Biochemistry* **41**, 3781–3789 (2002).
37. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
38. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).
39. Abrahams, J. P. & Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F₁ ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
40. Cowtan, K. Error estimation and bias correction in phase-improvement calculations. *Acta Crystallogr. D* **55**, 1555–1567 (1999).
41. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
42. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
43. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).
44. Tao, M. *et al.* Crystallization of the NADH-oxidizing domain of the Na⁺-translocating NADH:ubiquinone oxidoreductase from *Vibrio cholerae*. *Acta Crystallogr. F* **62**, 110–112 (2006).
45. Vohl, G. *et al.* Crystallization and preliminary analysis of the NqrA and NqrC subunits of the Na⁺-translocating NADH:ubiquinone oxidoreductase from *Vibrio cholerae*. *Acta Crystallogr. F* **70**, 987–992 (2014).
46. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66**, 22–25 (2010).
47. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
48. Hadd, J. J. *et al.* Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr. D* **68**, 381–390 (2012).
49. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
50. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
51. Chovancova, E. *et al.* CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLOS Comput. Biol.* **8**, e1002708 (2012).
52. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
53. The PyMOL Molecular Graphics System, v. 1.3r1 (Schrödinger, LLC, 2010).
54. Sääf, A., Johansson, M., Wallin, E. & von Heijne, G. Divergent evolution of membrane protein topology: the *Escherichia coli* RnfA and RnfE homologues. *Proc. Natl Acad. Sci. USA* **96**, 8540–8544 (1999).
55. Bogachev, A. V., Bertsova, Y. V., Bloch, D. A. & Verkhovsky, M. I. Thermodynamic properties of the redox centres of Na⁺-translocating NADH:quinone oxidoreductase. *Biochemistry* **45**, 3421–3428 (2006).
56. Bogachev, A. V. *et al.* Redox properties of the prosthetic groups of Na⁺-translocating NADH:quinone oxidoreductase. 1. Electron paramagnetic resonance study of the enzyme. *Biochemistry* **48**, 6291–6298 (2009).

NqrB

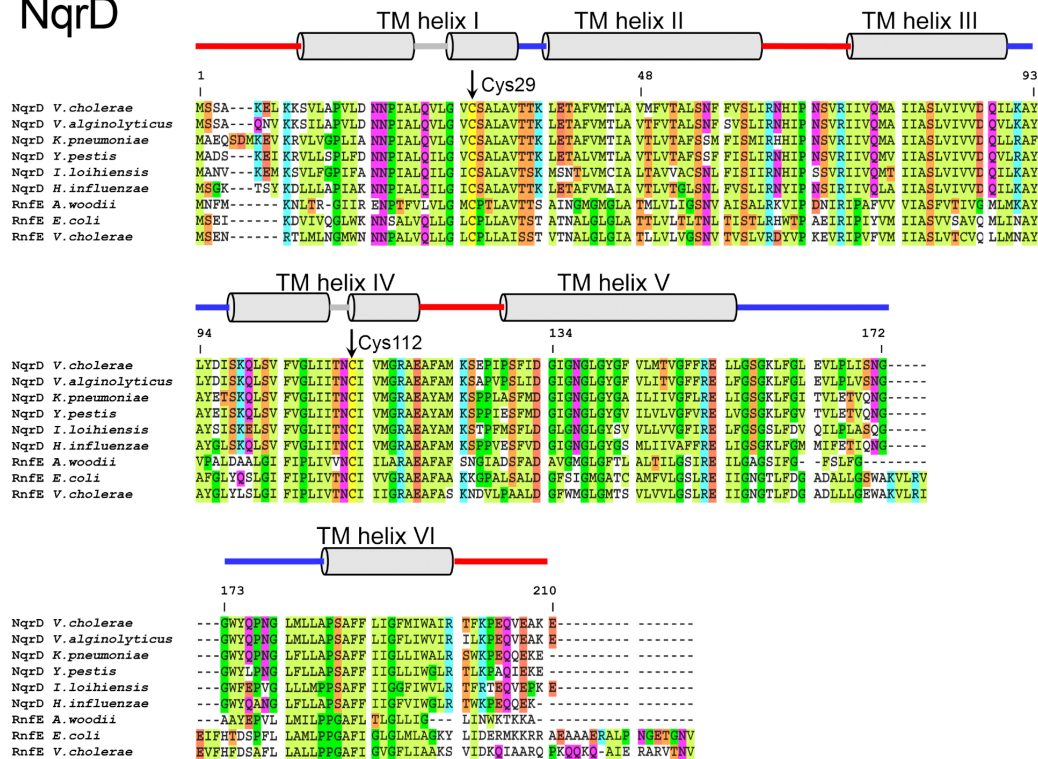


Extended Data Figure 1 | Sequence alignment of the integral membrane subunit NqrB from different organism with the corresponding subunits of the RNF complex. The localization of transmembrane helices is indicated by cylinders. Connecting loops located in the cytoplasm are shown in red,

connecting loops located in the periplasm in blue. Thr 236 covalently binding the FMN, and Asp 346 located in the proposed Na⁺ channel are indicated by arrows.

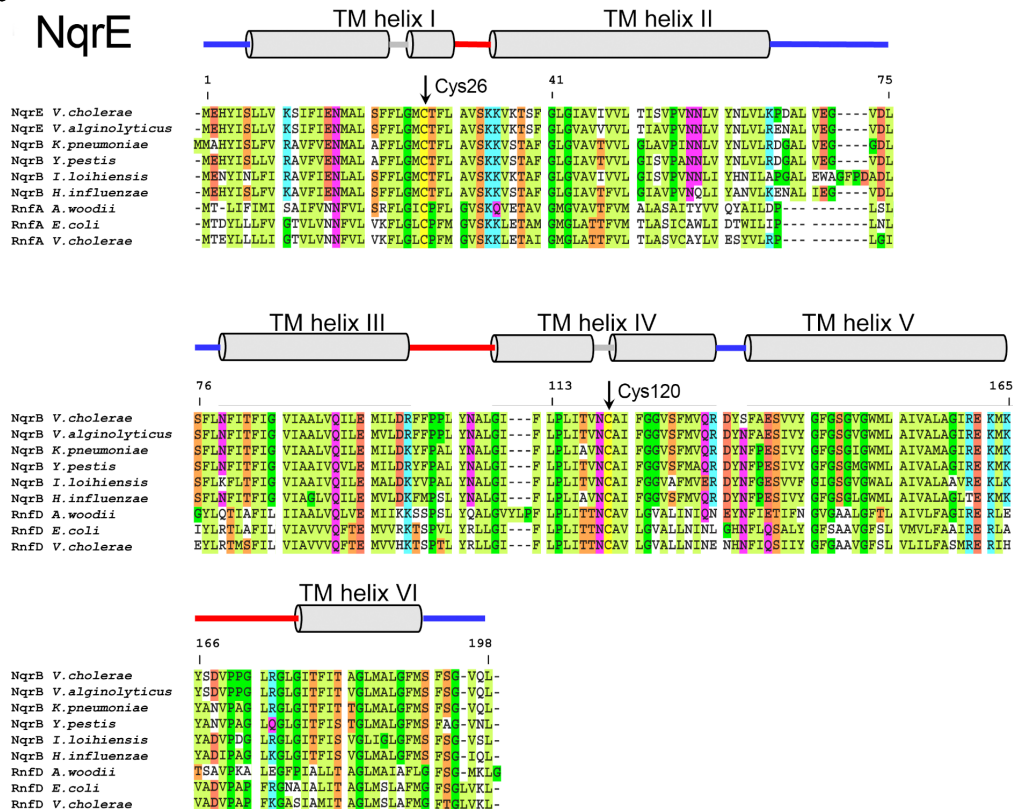
a

NqrD



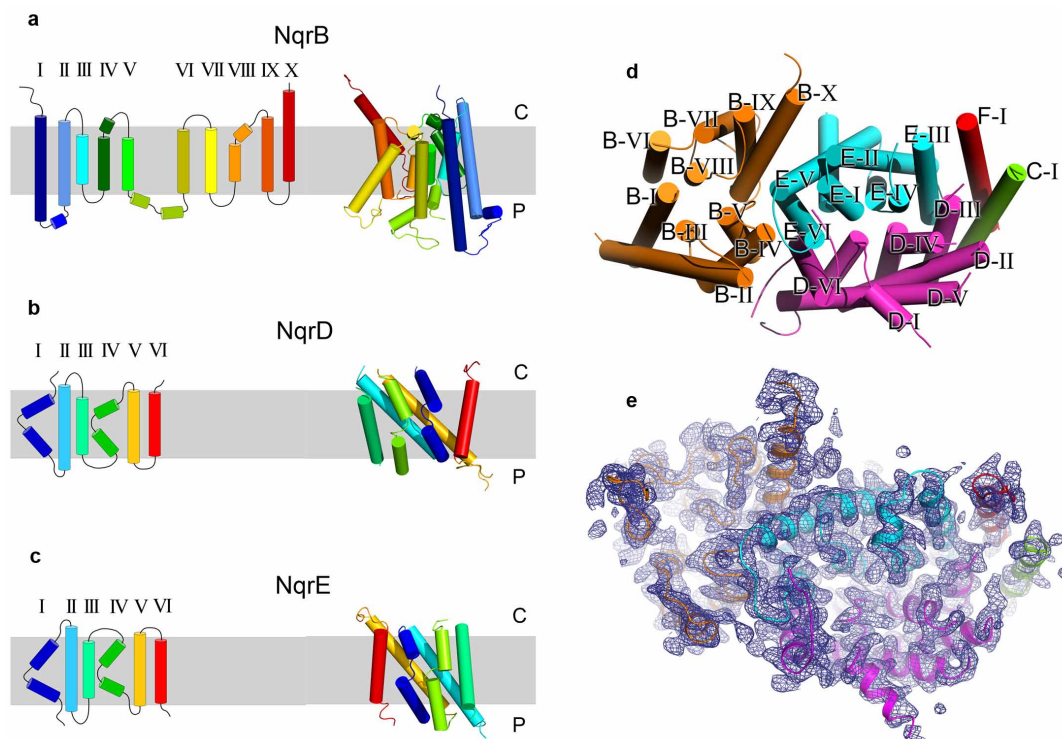
b

NqrE



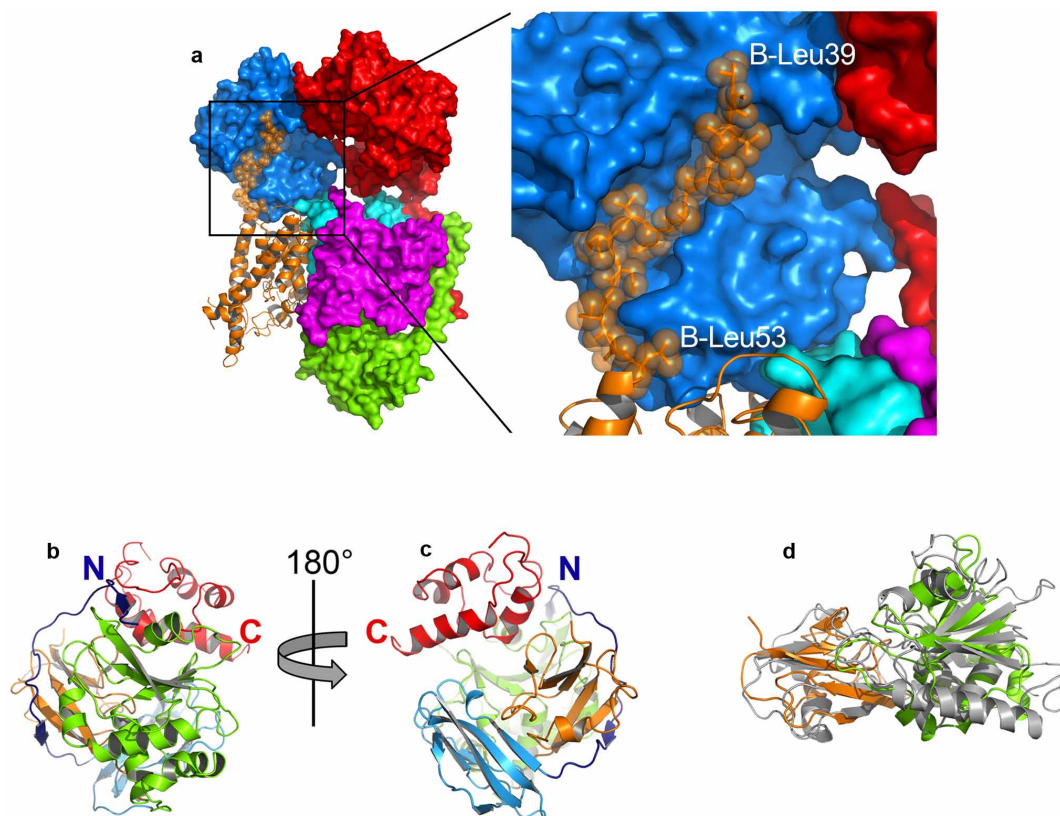
Extended Data Figure 2 | Sequence alignment of the integral membrane subunits NqrD and NqrE from different organism with the corresponding subunits of the RNF complex. a, b, The localization of transmembrane helices

is indicated by cylinders; connecting loops located in the cytoplasm are shown in red, connecting loops located in the periplasm in blue. Cys residues in NqrD (a) and NqrE (b) coordinating the Fe are indicated by arrows.



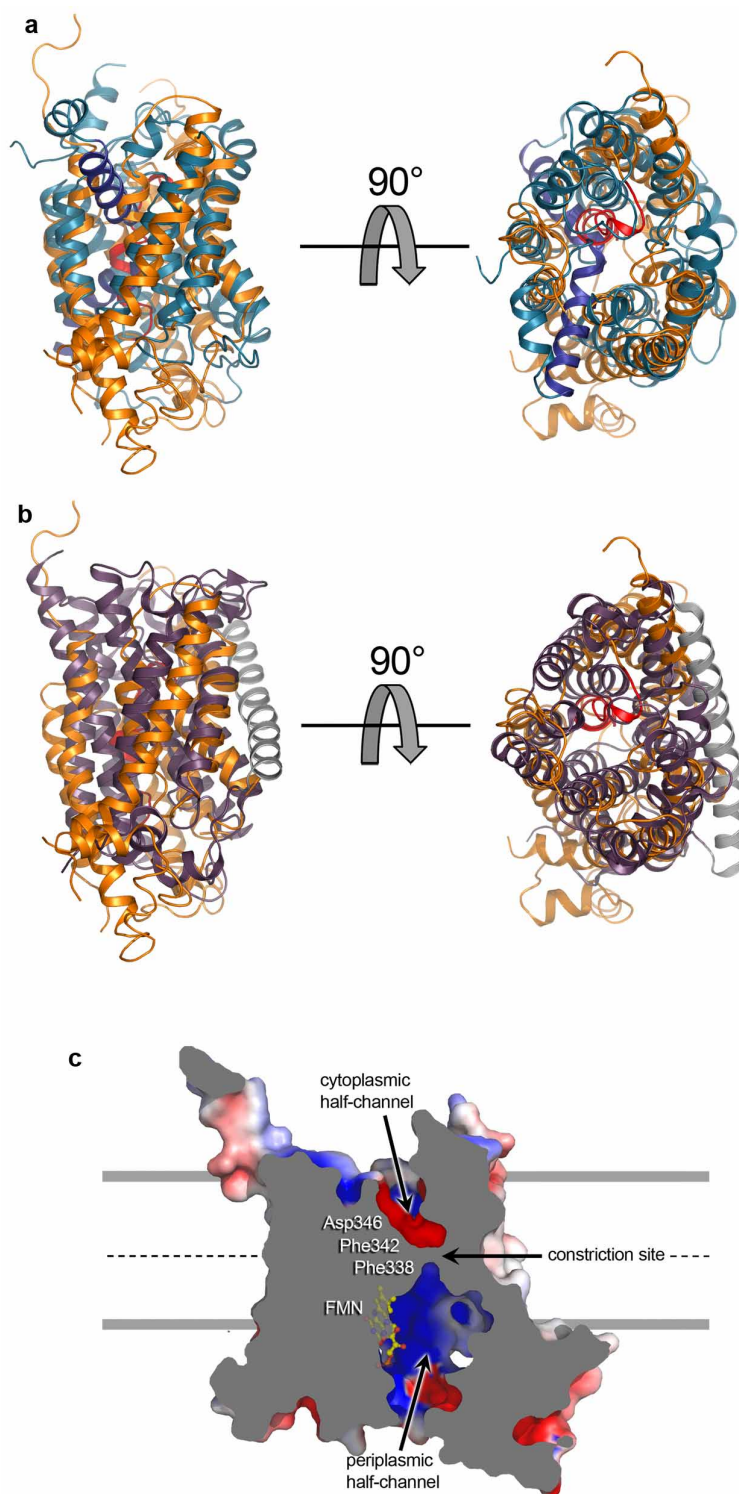
Extended Data Figure 3 | Topology of the transmembrane subunits NqrB, NqrD, and NqrE and arrangement of transmembrane helices. **a–c,** The schematic topology of the transmembrane helices of NqrB, NqrC and NqrD is shown on the left hand side and the corresponding structure on the right hand side. The membrane plane is indicated in grey and the cytoplasmic aspect is marked by C and the periplasmic aspect by P. **a,** NqrB contains ten transmembrane helices which can be divided into a N-terminal domain comprising helices I–V and a C-terminal domain comprising helices VI–X, which exhibit an inverted topology. Both domains are connected by a long periplasmic linker. The domains exhibit an inverted topology and align with an r.m.s.d. of 3.3 Å over 113 C α positions. **b, c,** NqrD and NqrE each comprise six helices exhibiting an inverted topology. Helix I and helix IV of both subunits are composed of two half helices. Such an inverted topology had been predicted based on the sequence information⁵⁴. **d,** Top view from the cytoplasmic side

onto the transmembrane helices of subunits NqrB, NqrC, NqrD, NqrE and NqrF. There are a total of 24 transmembrane helices. NqrD and NqrE form a central symmetrical unit. Subunit NqrB resides on one side of the NqrD–E unit whereas the single transmembrane helices from NqrC and NqrF reside on the opposed side. NqrB is closely attached to NqrE via helices V and VI from NqrE and IV, V, IX and X from NqrB, forming an interaction surface of 1,280 Å², whereas NqrD exhibits a much smaller contact area to NqrB via helices VI from NqrD and IV and V from NqrB, covering 335 Å². The transmembrane helices of NqrC and NqrF are close to each other but interact with different subunits: the transmembrane helix of NqrC forms contacts with helix III of NqrD, whereas the transmembrane helix of NqrF interacts with helix III of NqrE. **e,** Top view of the transmembrane part of Na⁺-NQR and 2F_o – F_c electron density displayed at a contour level of 1.0 σ . The map coefficients were sharpened by a B-factor of –80 Å².



Extended Data Figure 4 | Subunit NqrA. **a**, Interactions of NqrA with other subunits in the Na⁺-NQR complex. The subunits of Na⁺-NQR are shown in different colours: NqrA in blue, NqrB in orange, NqrC in green, NqrD in magenta, NqrE in cyan, and NqrF in red. Subunit B is shown as cartoon and all other subunits as surface representation. The C-terminal domain of NqrA located proximal to the membrane forms minor contacts with the integral membrane subunit NqrB via the NqrA residues 376–379 and 425–428, located in two short loops. A long N-terminal stretch of NqrB encompassing residues 39–53 lies in a groove of NqrA interacting over a total area of 820 Å² and anchoring NqrA to the membrane subunits. The residues shown as transparent van der Waals spheres fill almost the entire groove of NqrA. At the C terminus of NqrB, transmembrane helix 10 is elongated and protrudes into the cytoplasm, forming contacts with the C-terminal domain and the Rossmann-fold domain of NqrA, covering a total area of 430 Å². **b**, **c**, NqrA is composed of four domains, an N-terminal domain similar to a biotin carboxyl carrier

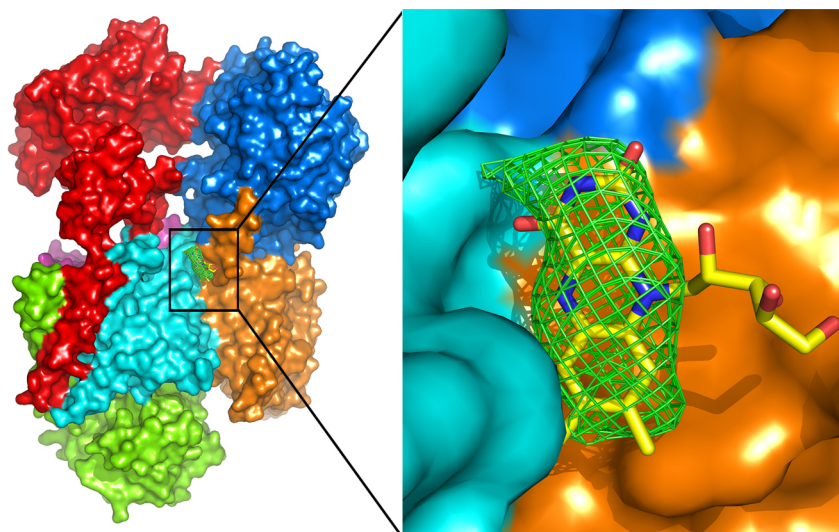
domain (blue, residues 28–100), a Rossmann-fold domain (green, residues 102–254), an ubiquitin-like domain (orange, residues 258–329), and a C-terminal helical domain (red, residues 376–446). The N-terminal residues 1–27 wrap around the Rossmann-fold domain and the ubiquitin-like domain and form two short β-strands that align with β-sheets of both domains, respectively. The C-terminal helical domain of NqrA shows similarity to a 2[4Fe–4S] cluster ferredoxin fold like for example, in fumarate reductase (PDB code 1KF6), but does not contain a FeS centre. Consistently, the Cys residues required for FeS coordination are not present in NqrA. **d**, Structural alignment of NqrA with Nqo1 from complex I (grey). The proteins align with an r.m.s.d. of 3.9 Å over 234 Cα positions. NqrA comprises a deep solvent-accessible cavity that is formed by residues of the Rossmann-fold domain and the ubiquitin-like domain that is large enough to accommodate ubiquinone. In case of Nqo1 of complex I the corresponding cavity harbours the isoalloxazine moiety of the FMN cofactor.



Extended Data Figure 5 | A putative Na⁺ channel in subunit NqrB.

a, b, Structural alignments of NqrB with urea transporter and ammonium transporter are shown. In NqrB the central helices I, III, VI and VIII form a membrane-spanning channel. Some backbone carbonyls, for example, from Val 161, Ile 164, Leu 168 from helix III deviate notably from the ideal geometry and point inwards the channel. Such a distortion indicates a putative involvement in Na⁺ coordination. **a,** The left hand side represents the side view and the right hand side the top view of NqrB (orange) aligned with bovine urea transporter (blue). Helix VIII of NqrB carrying residues forming the constriction is shown in red. The gating helices of urea transporter, which have no corresponding helices in NqrB, are depicted in dark blue. **b,** Structural alignment of NqrB (orange) with ammonium transporter from *Archaeoglobus fulgidus*. The outer helix of ammonium transporter that has no homologous

helix in NqrB is shown in grey. The high structural similarity of NqrB with urea and ammonium transporter shows that the subunit preserved the basic architecture of a transporter, but has acquired an additional and completely different function as a redox protein. These structural rearrangements in the periplasmic aspect of NqrB required to embed the FMN cofactor might have contributed to the closure of the channel. **c,** Cross section through NqrB. The surface is coloured according to the electrostatic surface potential. The cytoplasmic half channel exhibits a negative surface charge (red) whereas the periplasmic half channel is positively charged (blue). The localization of residues Phe 338, Phe 342 and Asp 346 is indicated. The constriction is located halfway through the membrane. The borders of the cytoplasmic membrane are indicated by grey lines.



Extended Data Figure 6 | Localization of riboflavin. A large patch of $F_o - F_c$ density was observed between NqrB (orange) and NqrE (cyan) and assigned to the riboflavin. The isoalloxazine moiety of riboflavin fits well into the $F_o - F_c$ density. Several interactions with the protein matrix can stabilize

the riboflavin. The flavin is stacked between the side chain of Val 399 and the CB, CG of Glu 402 of NqrB on one side (*Si* side) and the side chain of Phe 39 of NqrE on the opposed side (*Re*-side). Moreover, the imidazole of His 398 of NqrB on the *Si*-side can form a hydrogen bond to N5 of isoalloxazine.

Extended Data Table 1 | Data collection, phasing and refinement statistics

	Native	Ta ₆ Br ₁₂	Se Met	Fe
Data collection				
X-ray source	SLS, X06SA	SLS, X06SA	SLS, X06SA	SLS, X06SA
Wavelength (Å)	0.90	1.225	0.9797 (peak)	1.736 (peak)
Total rotation range	720°	1200°	4720°	1600°
Space group	<i>P</i> 2 ₁	<i>P</i> 2 ₁	<i>P</i> 2 ₁	<i>P</i> 2 ₁
Unit-cell parameters				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	95.3, 143.4, 104.5	96.7, 151.3, 106.1	96.2, 150.8, 108.2	95.7, 150.3, 107.5
α , β , γ (°)	90, 110.9, 90	90, 111.5, 90	90, 111.8, 90	90, 111.6, 90
Resolution (Å)	50-3.5 (3.7-3.5)	50-4.2 (4.4-4.2)	50-3.8 (4.0-3.8)	50-4.0 (4.2-4.0)
<i>R</i> _{merge} (%)	17.1 (563.8)	7.8 (174.5)	12.6 (383.4)	6.7 (135.1)
<i>R</i> _{meas} (%)	17.8 (584.9)	8.2 (182.5)	12.8 (387.5)	6.9 (139.9)
<i>R</i> _{pim} (%)	4.8 (156.4)	2.4 (53.7)	1.8 (56.8)	1.7 (36.0)
<i>CC</i> _{1/2}	0.999 (0.411)	1.000 (0.568)	1.000 (0.362)	1.000 (0.496)
<i>I</i> / σ (<i>I</i>)	10.9 (0.86)	11.9 (1.35)	18.4 (1.4)	14.4 (2.0)
Completeness (%)	99.7 (99.8)	99.4 (99.4)	99.5 (99.5)	99.5 (98.9)
Multiplicity	13.7 (14.0)	11.7 (11.6)	46.6 (46.6)	15.5 (15.1)
No. molecules per ASU	1	1	1	1
Solvent content (%)	59.9	62.9	63.2	62.7
Refinement				
Resolution (Å)	50-3.5 (3.7-3.5)	50-4.2 (4.4-4.2)	50-3.8 (4.0-3.8)	50-4.0 (4.2-4.0)
No. of reflections	452,674 (70,668)	477,442 (62,286)	2,588,229 (370,515)	726,668 (95,391)
No. of unique reflections	33,122 (5,032)	40,704 (5,337)	55,473 (7,935)	46,881 (6,336)
<i>R</i> _{work} (%)	24.6 (35.2)			
<i>R</i> _{free} (%)	27.6 (38.1)			
No. atoms				
Protein	13,607			
Ligand/ion	13,461			
	146			
Mean B-factors (Å ²)				
Protein	167.7			
Ligand/ion	167.5			
	183.7			
R.m.s. deviations				
Bond lengths (Å)	0.011			
Bond angles (°)	1.12			
MolProbity overall score				
Clashscore	2.99			
	14.14			
Ramachandran plot (%)				
(favoured/allowed/outliers)	(90.2/7.0/2.8)			

Extended Data Table 2 | Fe anomalous map peak heights

Crystal No.	Wavelength (Å)	Anomalous peak height at position of [2Fe-2S] _{NqrF}	Anomalous peak height at position of Fe _{NqrD-E}
1	1.7268 (remote high)	9.1	11.8
1	1.7360 (peak)	8.4	13.2
1	1.7428 (inflection)	8.4	13.7
1	1.7463 (remote low)	3.0	2.7
2	1.7287 (remote high)	5.3	6.0
2	1.7360 (peak)	7.6	9.3
2	1.7428 (inflection)	4.7	3.6
2	1.7463 (remote low)	0	0

To prove the identity of Fe causing the strong peak in the anomalous map localized in subunits NqrD-E, we recorded data sets at different wavelengths close to the K-absorption edge of Fe. The wavelengths were chosen according to a fluorescence scan around the Fe K-edge. The resulting anomalous peak heights at the different wavelengths are similar for both, the [2Fe-2S]_{NqrF} site and the Fe_{NqrD-E} site. The maximal peak height is observed at a wavelength corresponding to the peak of absorption. Changing the wavelength towards lower energy results in a sharp drop and no anomalous signal is observed at a wavelength where Fe does not absorb. Multiple fluorescence scans on different crystals revealed no further presence of any metal ion in Na⁺-NQR. The similar peak height at the [2Fe-2S]_{NqrF} site and Fe_{NqrD-E} site might be attributed to conformational flexibility of subunit NqrF as indicated by higher *B*-factors compared to the integral membrane subunits. Such flexibility will result in a weaker anomalous signal from the [2Fe-2S]_{NqrF} centre.

Extended Data Table 3 | Redox cofactor distances

Redox cofactor pair	Edge-to-edge (Å)	Centre-to-centre (Å) Atom ID	E^0 (mv) *
FAD _{NqrF} / [2Fe-2S] _{NqrF}	9.8	16.0 (FAD-N5 / [2Fe-2S]-Fe1)	-200 / -275
[2Fe-2S] _{NqrF} / Fe _{NqrDIE}	33.2	36.4 ([2Fe-2S]-Fe1 / Fe _{NqrDIE})	-200 / n.d.
[2Fe-2S] _{NqrF} / Riboflavin	29.4	34.6 ([2Fe-2S]-Fe1 / Riboflavin-N5)	-200 / -10
Fe _{NqrDIE} / FMN _{NqrC}	7.9	13.2 (Fe _{NqrDIE} / FMN _{NqrC} -N5)	n.d. / -190
Fe _{NqrDIE} / Riboflavin	30.8	32.8 (Fe _{NqrDIE} / Riboflavin-N5)	n.d. / -10
Fe _{NqrDIE} / FMN _{NqrB}	29.6	32.9 (Fe _{NqrDIE} / FMN _{NqrB} -N5)	n.d. / -130
FMN _{NqrC} / FMN _{NqrB}	21.4	27.2 (FMN _{NqrC} -N5 / FMN _{NqrB} -N5)	-190 / -130
FMN _{NqrB} / Riboflavin	29.3	32.7 (FMN _{NqrB} -N5 / Riboflavin-N5)	-190 / -130

* The values are taken from refs 55 and 56.

Extended Data Table 4 | r.m.s.d. deviations between subunits in NQR complex and the structures of the individual subunits

Na ⁺ -NQR / subunit pair	No. C α	R.M.S.D. (Å)
Na ⁺ -NQR / NqrA ₁₋₃₇₇ SG C222 ₁ (PDB 4U9O)	298	0.82
Na ⁺ -NQR / NqrA ₁₋₃₇₇ SG P2 ₁ (PDB 4U9Q)	293	3.86
Na ⁺ -NQR / NqrC ₃₃₋₂₅₇ (PDB 4U9S)	218	1.01
Na ⁺ -NQR / NqrF ₁₂₉₋₄₀₈ (PDB 4U9U)	279	0.69

The high-resolution structures of the individual subunits deviate only to a minor extent from the corresponding parts in the entire complex. Largest differences are observed for NqrA₁₋₃₇₇, which was crystallized in two different space groups. NqrA₁₋₃₇₇, crystallized in space group C222₁, displays a r.m.s.d. of 0.82 Å, whereas NqrA₁₋₃₇₇, crystallized in space group P2₁, exhibits a r.m.s.d. of 3.86 Å. This deviation originates mainly from a different conformation in the NqrA region 213–235 and from different conformations of two loops encompassing residues 155–161 and 292–297, respectively. In the structure of the entire NqrA in the NQR complex this region is in contact with the C-terminal domain of NqrA and forms a short α -helix (residues 219–229), whereas this region forms an extended loop stabilized by crystal contacts in NqrA₁₋₃₇₇ in space group P2₁. In crystals of NqrA₁₋₃₇₇ in space group C222₁ this region was completely unordered and could not be resolved in the electron density.

Stellar feedback as the origin of an extended molecular outflow in a starburst galaxy

J. E. Geach¹, R. C. Hickox², A. M. Diamond-Stanic³, M. Krips⁴, G. H. Rudnick^{5,6}, C. A. Tremonti³, P. H. Sell⁷, A. L. Coil⁸ & J. Moustakas⁹

Recent observations have revealed that starburst galaxies can drive molecular gas outflows through stellar radiation pressure^{1,2}. Molecular gas is the phase of the interstellar medium from which stars form, so these outflows curtail stellar mass growth in galaxies. Previously known outflows, however, involve small fractions of the total molecular gas content and have typical scales of less than a kiloparsec^{1,2}. In at least some cases, input from active galactic nuclei is dynamically important^{2,3}, so pure stellar feedback (the momentum return into the interstellar medium) has been considered incapable of rapidly terminating star formation on galactic scales. Molecular gas has been detected outside the galactic plane of the archetypal starburst galaxy M82 (refs 4 and 5), but so far there has been no evidence that starbursts can propel substantial quantities of cold molecular gas to the same galactocentric radius (about 10 kiloparsecs) as the warmer gas that has been traced by metal ion absorbers in the circumgalactic medium^{6,7}. Here we report observations of molecular gas in a compact (effective radius 100 parsecs) massive starburst galaxy at redshift 0.7, which is known to drive a fast outflow of ionized gas⁸. We find that 35 per cent of the total molecular gas extends approximately 10 kiloparsecs, and one-third of this extended gas has a velocity of up to 1,000 kilometres per second. The kinetic energy associated with this high-velocity component is consistent with the momentum flux available from stellar radiation pressure^{9–12}. This demonstrates that nuclear bursts of star formation are capable of ejecting large amounts of cold gas from the central regions of galaxies, thereby strongly affecting their evolution by truncating star formation and redistributing matter^{13,14}.

SDSS J0905+57 (redshift $z = 0.712$) is a compact starburst galaxy with emission-line properties consistent with a star-forming galaxy and no observational evidence of a strong hot dust continuum in the mid-infrared part of the spectrum, indicating no significant black hole accretion activity^{8,15}. The galaxy is driving a wind with one of the highest velocities known for any star-forming galaxy, with the interstellar absorption lines of Ca II, Fe II and Mg II blueshifted by 2,500 km s⁻¹ with respect to the Balmer stellar absorption lines. The total infrared luminosity is $L_{\text{IR}} \approx 1.2 \times 10^{39}$ W, corresponding to a star formation rate (SFR) of 260 M_{\odot} yr⁻¹, where M_{\odot} is the mass of the Sun. Hubble Space Telescope observations reveal that SDSS J0905+57 is extremely compact in the rest-frame V band (475 nm), with an effective radius of $r_e = 94$ pc (comparable to the size of 30 Doradus). This implies an SFR density of $\Sigma_{\text{SFR}} \approx 4,700 M_{\odot}$ yr⁻¹ kpc⁻². The compact nature of the galaxy and the high density of central star formation suggest that SDSS J0905+57 is likely to be at the final stage of a major merger¹⁵ and is the progenitor of an elliptical galaxy.

We observed SDSS J0905+57 with the Institut de Radioastronomie Millimétrique Plateau de Bure Interferometer in the 2 mm band with receivers tuned to the frequency of the redshifted CO(2–1) emission line at $z = 0.712$ (134 GHz). At temperatures of the order of 10 K, the

carbon monoxide $J = 2 \rightarrow 1$ rotational transition is excited at a critical density of $n_{\text{crit}} \approx 10^4$ cm⁻³ and is a good tracer of the bulk of the cold molecular gas reservoir. The spectrum (Fig. 1) reveals a detection of the CO(2–1) emission line at observed frequency $\nu_{\text{obs}} = 134.666$ GHz, corresponding to $z_{\text{CO}} = 0.712$, which is consistent with the stellar redshift, with the full width at half maximum (FWHM) approximately 200 km s⁻¹. We refer to this as the ‘core’ line. The spectrum also reveals CO emission in a broad wing extending up to 1,000 km s⁻¹ from the core line. When averaged over the velocity range $\Delta V = 200$ –1,000 km s⁻¹, the emission is significant and peaks $1.2 \pm 0.3''$ or 8 ± 2 kpc from the core line with a flux density of $S = 0.43 \pm 0.09$ mJy (Figs 2 and 3). We interpret these observations as evidence that molecular gas is being driven out of the galaxy through stellar feedback processes.

The core CO line emission is also marginally resolved beyond the 3'' beam when averaged over the full core linewidth (full width at zero intensity, FWZI) ($\Delta V = 400$ km s⁻¹, Fig. 2). To confirm this, and to measure the size of the extended CO emission, we examine the u – v plane visibilities to evaluate the average signal amplitude as a function of baseline separation (Methods). The data are inconsistent with a flat profile that would indicate an unresolved source, but are better fitted by a combination of a point source and circular Gaussian profile with a half

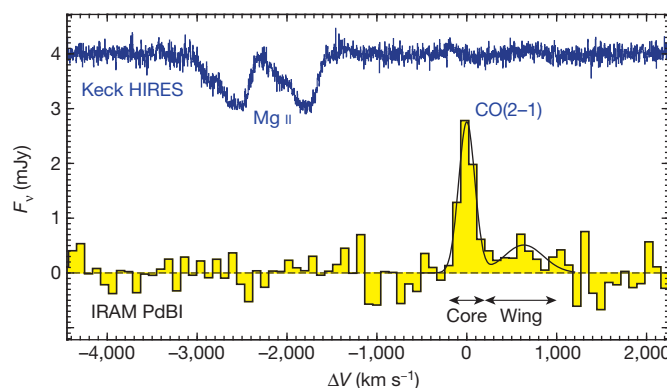


Figure 1 | The 2 mm spectrum of SDSS J0905+57 obtained with the IRAM Plateau de Bure Interferometer (PdBI). The observed flux density F_v of the CO(2–1) line is fitted with a Gaussian profile of width 200 km s⁻¹ (FWHM) and peaks at a redshift consistent with the stellar absorption lines. There is evidence for significant CO emission in a high-velocity wing that extends up to 1,000 km s⁻¹ from the core line, which could indicate a high-velocity outflowing molecular gas component; we jointly fitted this with a Gaussian profile. SDSS J0905+57 is also driving a high-velocity outflow of ionized gas, as revealed by the strongly blue-shifted Mg II doublet (at wavelengths $\lambda = 2,796$ and $\lambda = 2,803$ Å) absorption observed in the Keck High Resolution Echelle Spectrometer (HIRES) rest-frame ultraviolet spectrum, shown here on the same velocity (but arbitrary flux) scale, relative to the 2,796 Å line.

¹Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, Hertfordshire AL10 9AB, UK. ²Department of Physics and Astronomy, Dartmouth College, Hanover, New Hampshire 03755, USA. ³Department of Astronomy, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. ⁴Institut de Radioastronomie Millimétrique, 300 rue de la Piscine, F-38406 Saint Martin d'Hères, France. ⁵Department of Physics and Astronomy, University of Kansas, Lawrence, Kansas 66045, USA. ⁶Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany. ⁷Department of Physics, Texas Tech University, Lubbock, Texas 79409-1051, USA. ⁸Center for Astrophysics and Space Sciences, University of California, San Diego, La Jolla, California 92093, USA. ⁹Department of Physics and Astronomy, Siena College, 515 Loudon Road, Loudonville, New York 12211, USA.

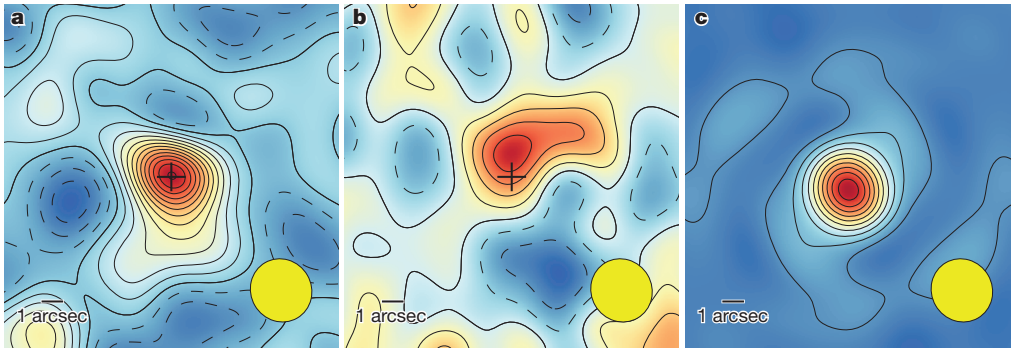


Figure 2 | Maps of carbon monoxide emission. **a, b,** Cleaned CO(2–1) line maps averaged over the FWZI core line ($\pm 200 \text{ km s}^{-1}$) (**a**) and the wing emission over $200\text{--}1,000 \text{ km s}^{-1}$ (**b**). The colour scales are a linear mapping of intensity over the 0.25%-to-99.75% percentile range in each image. Contours are spaced at multiples of the root mean squared noise level in the maps, with dashed contours tracing negative deviations. **b,** The wing emission peaks

$1.2 \pm 0.3''$ ($8 \pm 2 \text{ kpc}$) from the core line with a peak flux of $0.43 \pm 0.09 \text{ mJy}$ (4.8σ). **c,** The shape of the dirty beam, with contours at 90% (innermost) to 10% (outermost) of the peak flux at 10% intervals. The yellow ellipse in all panels represents the FWHM of the dirty beam, approximately $3''$. The crosses mark the position of the peak of the emission in **a**.

power radius of $1.6^{+0.8}_{-0.4}$ arcseconds. A point-source-only model can be ruled out at the 4.7σ level. The angular size of this extended component corresponds to a radius of 12^{+6}_{-3} kpc in physical projection, 130 times larger than the rest-frame V-band effective radius (Fig. 3). This extended low-velocity CO emission could also be associated with feedback processes (for example, previously ejected gas), but we cannot rule out the hypothesis that it represents molecular gas ejected from disks during previous stages of the merger.

We assume that the unresolved CO component is associated with dense gas still actively forming stars. The mass of this active component is estimated as $M_{\text{H}_2} = (3.1 \pm 0.6) \times 10^9 M_{\odot}$, assuming thermalized CO $J = 2 \rightarrow 1$ emission and $\alpha = 0.8 M_{\odot} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$, where $\alpha L'_{\text{CO}} = M_{\text{H}_2}$ (L'_{CO} is the CO luminosity in units of $\text{K km s}^{-1} \text{ pc}^2$) and α is the CO-to- H_2 conversion factor appropriate for the conditions in the nuclear regions of ultraluminous infrared galaxies¹⁶. The infrared-to-CO luminosity ratio $L_{\text{IR}}/L'_{\text{CO}} \approx 800 L_{\odot} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$ is close to the upper limit predicted for star formation limited by radiation pressure¹⁰. The masses in the extended low-velocity ($|\Delta V| < 200 \text{ km s}^{-1}$) and high-velocity ($\Delta V = 200\text{--}1,000 \text{ km s}^{-1}$) wing components are $M_{\text{H}_2} = (1.1 \pm 0.5) \times 10^9 M_{\odot}$ and $M_{\text{H}_2} = (0.6 \pm 0.2) \times 10^9 M_{\odot}$ respectively. Combined,

the extended CO emission represents approximately 35% of the total gas mass. An uncertainty here is the choice of α ; we assume a conservative value of $\alpha = 0.34 M_{\odot} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$ for the extended components, which assumes local thermodynamic equilibrium and is applicable to the optically thin case of turbulent gas associated with a wind, assuming an CO/ H_2 abundance of 10^{-4} and typical excitation temperature of 30 K (ref. 17).

If the extended CO-emitting gas forms a foreground screen, then the total hydrogen atom column density N_{H} can be inferred in the limit where the column is dominated by molecular gas. As with the molecular mass estimate, one must adopt a value for the ‘X-factor’ that relates CO emission to molecular hydrogen column density $X_{\text{CO}} = N_{\text{H}_2}/W_{\text{CO}}$, with N_{H_2} in units of cm^{-2} and the integrated line intensity W_{CO} in units of K km s^{-1} . We assume $X_{\text{CO}} = 1.6 \times 10^{19} \text{ cm}^{-2} (\text{K km s}^{-1})^{-1}$, for the same assumptions as for α for the extended components described above, which yields $N_{\text{H}} \approx 2 \times 10^{20} \text{ cm}^{-2}$. Independently, the extinction of the stellar and nebular emission can be estimated from the relative intensities of the Balmer lines, which indicate an extinction of $A_V \approx 0.5 \text{ mag}$ to the young, compact stellar population (assuming Milky Way abundances¹⁸). This corresponds to a column density of $N_{\text{H}} \approx 10^{21} \text{ cm}^{-2}$, in reasonable agreement with the value estimated from the cold gas. A plausible scenario is that an outflow, or series of outflows, launched from the nuclear starburst has purged the interstellar medium of the stellar bulge, sweeping cold gas and dust into the halo. This ‘blow out’ phase will rapidly truncate star formation in the bulge on a timescale comparable to the dynamical time, exposing a bright shell of young stars around the nuclear starburst, which will consume the remaining molecular gas within 10 million years given the gas supply and consumption rate.

For cold gas to be driven to large galactocentric distances, the wind driving mechanism must be favourable to the survival of cold clouds⁹. Originally it was thought that cold material would be ejected along with the hot gas associated with the explosions of supernovae¹⁹, but cold clouds entrained in such outflows are predicted to be destroyed on timescales of millions of years and quickly incorporated into the hot flow²⁰. Alternatively, stellar radiation pressure on dust grains can accelerate cold gas several million years before the first cluster supernovae explode, without subjecting it to the deleterious effects of a hot conductive atmosphere⁹. This mechanism has already been shown to be the most likely driver in local starbursts exhibiting subkiloparsec molecular outflows². After exiting the galaxy, the cold gas interacts with the potentially hot halo atmosphere. Although the hydrodynamic interaction between a hot (10^6 K) atmosphere and high-velocity cold (10 K) to cool (10^4 K) clouds is complex²¹, these observations suggest that ejected molecular gas can survive in such an environment for timescales of at least approximately 10 million years.

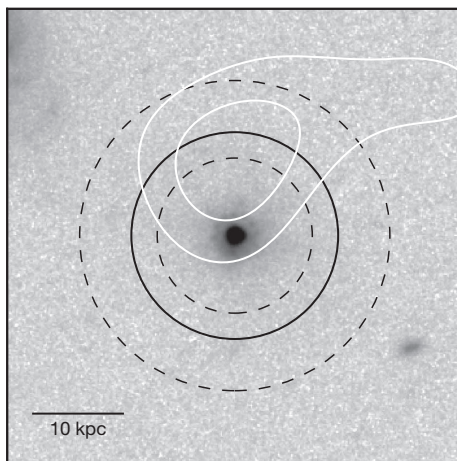


Figure 3 | Optical image of SDSS J0905+57 from the Hubble Space Telescope. The optical image reveals the rest-frame V-band morphology of the target obtained with the Wide Field Camera 3 Ultraviolet Imaging Spectrograph (F814W filter). The galaxy has an effective radius of 100 pc for the rest-frame V-band light. White contours show the 3σ to 4σ levels (4σ is the innermost and highest contour) for CO emission averaged over $200\text{--}1,000 \text{ km s}^{-1}$. The black circle shows the CO half-power size when averaged over the FWZI of the core line. Dashed lines show the 1σ confidence bounds for the half-power size.

If the maximum velocity ($V \approx 1,000 \text{ km s}^{-1}$) in the redshifted wing is representative of the deprojected velocity of the outflow²², then the extended CO emitting gas carries kinetic power $P_K = (2.6 \pm 0.8) \times 10^{36} \text{ W}$. The mass outflow rate is $\dot{M}_{\text{H}_2} = 80 \pm 25 M_\odot \text{ yr}^{-1}$, implying a ‘mass-loading’ factor of $\dot{M}_{\text{H}_2}/\text{SFR} \approx 30\%$. In a momentum-driven wind, the mass loading factor scales inversely with outflow velocity²³, and our results are roughly consistent with an extrapolation of the linear correlation between $\dot{M}_{\text{H}_2}/\text{SFR}$ and v that has been found for local pure starbursts². The rate of momentum input available from stellar radiation pressure in the single scattering limit is $L/c \approx 4 \times 10^{30} \text{ N}$ (ref. 24), where L is the bolometric luminosity, but note that this will be larger when the medium is optically thick to far-infrared photons, scaling with the optical depth τ_{FIR} (ref. 9). The outflow momentum flux is $v\dot{M}_{\text{H}_2} = (4.8 \pm 1.9) \times 10^{30} \text{ N}$, implying that the energy of the outflow is compatible with what is available from star formation alone. Material travelling at $1,000 \text{ km s}^{-1}$ takes less than 10 million years to reach the measured radial extent ($r = 8 \pm 2 \text{ kpc}$) of the outflow. Again, this estimate is uncertain owing to projection effects, but the timescale is in broad agreement with the age of the young stellar population. Fitting of the rest-frame ultraviolet/optical spectra reveals that 90% of the stellar luminosity is contributed by a population of age 6 million years or younger. We cannot rule out the possibility that the outflow was launched by an active galactic nucleus (AGN) that has since ‘switched off’, but the current observations indicate that the outflow is compatible with pure stellar feedback.

A key goal in galaxy evolution studies has been to understand the coupling between various forms of energy and momentum injection and the cold interstellar medium, as well as their relative efficacies as feedback channels. The molecular outflow in SDSS J0905+57 at $z = 0.7$ is extended on a much larger scale than has been previously observed in local starburst galaxies exhibiting molecular winds^{1,2}. In part, this could be related to the ultra-compact morphology and extreme nature of the system, which has a star-formation-rate density that is orders of magnitude larger than the local systems. In local galaxies, where small-scale radiatively driven molecular winds have been observed, only a few per cent of the total gas reservoir is involved in the outflow^{1,2}, whereas in SDSS J0905+57 up to a third of the total molecular gas reservoir appears to have been ejected. These observations are evidence that pure stellar feedback can affect the evolution of a galaxy as a whole on short (millions of years) timescales by directly removing the dense material required for star formation, and is therefore competitive with AGN feedback^{25–27} as a mechanism for regulating stellar mass growth and redistributing baryons in massive galaxies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 August; accepted 29 October 2014.

1. Bolatto, A. D. *et al.* Suppression of star formation in the galaxy NGC253 by a starburst-driven molecular wind. *Nature* **499**, 450–453 (2013).
2. Ciccone, C. *et al.* Massive molecular outflows and evidence for AGN feedback from CO observations. *Astron. Astrophys.* **562**, 21 (2014).
3. Tadhunter, C., Morganti, R., Rose, M., Oonk, J. B. R. & Oosterloo, T. Jet acceleration of the fast molecular outflows in the Seyfert galaxy IC 5063. *Nature* **511**, 440–443 (2014).
4. Walter, F., Weiss, A. & Scoville, N. Molecular gas in M82: resolving the outflow and streamers. *Astrophys. J.* **580**, L21–L25 (2002).
5. Salak, D. *et al.* ($J = 1-0$) observations of the starburst galaxy M82. *Publ. Astron. Soc. Jpn* **65**, 66 (2013).
6. Rubin, K. H. R. *et al.* Low-ionization line emission from a starburst galaxy: a new probe of a galactic-scale outflow. *Astrophys. J.* **728**, 55–61 (2011).

7. Martin, C. L. *et al.* Scattered emission from $z \sim 1$ galactic outflows. *Astrophys. J.* **770**, 41–60 (2013).
8. Diamond-Stanic, A. M. *et al.* High-velocity outflows without AGN feedback: Eddington-limited star formation in compact massive galaxies. *Astrophys. J.* **755**, L26–L30 (2012).
9. Murray, N., Ménard, B. & Thompson, T. A. Radiation pressure from massive star clusters as a launching mechanism for super-galactic winds. *Astrophys. J.* **735**, 66–78 (2011).
10. Andrews, B. H. & Thompson, T. A. Assessing radiation pressure as a feedback mechanism in star-forming galaxies. *Astrophys. J.* **727**, 97–108 (2011).
11. Hopkins, P. F., Quataert, E. & Murray, N. Stellar feedback in galaxies and the origin of galaxy-scale winds. *Mon. Not. R. Astron. Soc.* **421**, 3522–3537 (2012).
12. Thompson, T. A., Fabian, A. C., Quataert, E. & Murray, N. Dynamics of dusty radiation pressure driven shells: fast outflows from galaxies, star clusters, massive stars, and AGN. Preprint at <http://arxiv.org/abs/1406.5206> (2014).
13. Hopkins, P. F., Murray, N., Quataert, E. & Thompson, T. A. A maximum stellar surface density in dense stellar systems. *Mon. Not. R. Astron. Soc.* **401**, L19–L23 (2010).
14. Pontzen, A. & Governato, F. How supernova feedback turns dark matter cusps into cores. *Mon. Not. R. Astron. Soc.* **421**, 3464–3471 (2012).
15. Sell, P. H. *et al.* Massive compact galaxies with high-velocity outflows: morphological analysis and constraints on AGN activity. *Mon. Not. R. Astron. Soc.* **441**, 3417–3443 (2014).
16. Solomon, P. M., Downes, D., Radford, S. J. E. & Barrett, J. W. The molecular interstellar medium in ultraluminous infrared galaxies. *Astrophys. J.* **478**, 144–161 (1997).
17. Bolatto, A. D., Wolfire, M. & Leroy, A. K. The CO-to-H₂ conversion factor. *Annu. Rev. Astron. Astrophys.* **51**, 207–268 (2013).
18. Shetty, R., Glover, S. C., Dullemond, C. P. & Klessen, R. S. Modelling CO emission—I. CO as a column density tracer and the X factor in molecular clouds. *Mon. Not. R. Astron. Soc.* **412**, 1686–1700 (2011).
19. Chevalier, R. A. & Clegg, A. W. Wind from a starburst galaxy nucleus. *Nature* **317**, 44–45 (1985).
20. Cooper, J. L., Bicknell, G. V., Sutherland, R. S. & Bland-Hawthorn, J. Three-dimensional simulations of a starburst-driven galactic wind. *Astrophys. J.* **674**, 157–171 (2008).
21. Narayanan, D. *et al.* The role of galactic winds on molecular gas emission from galaxy mergers. *Astrophys. J.* **176** (suppl.), 331–354 (2008).
22. Maiolino, R. *et al.* Evidence of strong quasar feedback in the early Universe. *Mon. Not. R. Astron. Soc.* **425**, L66–L70 (2012).
23. Murray, N., Quataert, E. & Thompson, T. A. On the maximum luminosity of galaxies and their central black holes: feedback from momentum-driven winds. *Astrophys. J.* **618**, 569–585 (2005).
24. Velleux, S., Cecil, G. & Bland-Hawthorn, J. Galactic winds. *Annu. Rev. Astron. Astrophys.* **43**, 769–826 (2005).
25. Granato, G. L., De Zotti, G., Silva, L., Bressan, A. & Danese, L. A physical model for the coevolution of QSOs and their spheroidal hosts. *Astrophys. J.* **600**, 580–594 (2004).
26. Croton, D. J. *et al.* The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *Mon. Not. R. Astron. Soc.* **365**, 11–28 (2006).
27. Bower, R. G. *et al.* Breaking the hierarchy of galaxy formation. *Mon. Not. R. Astron. Soc.* **370**, 645–655 (2006).

Acknowledgements J.E.G. acknowledges support from the Royal Society through a University Research Fellowship. A.M.D.-S. acknowledges support from the Grainger Foundation. G.H.R. acknowledges the support of the Alexander von Humboldt Foundation and the hospitality of the Max Planck Institute for Astronomy. A.L.C. acknowledges funding from NSF CAREER grant AST-1055081. We thank E. Brinks, N. Murray, D. Narayanan, R. Neri and F. Walter for advice and discussions. This work is based on observations carried out with the IRAM Plateau de Bure Interferometer. IRAM is supported by INSU/CNRS (France), MPG (Germany) and IGN (Spain). Some of the data presented herein were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W. M. Keck Foundation.

Author Contributions J.E.G. and R.C.H. led the original IRAM observation proposals. All authors assisted with the data analysis and writing of the manuscript. J.E.G. and M.K. led the IRAM data reduction and analysis. J.M., A.M.D.-S. and C.A.T. led the analysis of the Keck spectroscopy and P.H.S. led the morphological analysis of the Hubble Space Telescope imaging.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.E.G. (j.geach@herts.ac.uk).

METHODS

Target information. The target SDSS J0905+57 at 09 h 05 min 23.7 s, +57° 59′ 12.6″ ($z = 0.712$) is a galaxy with the fastest Mg II outflow velocity of a larger sample of similar galaxies originally selected on the basis of post-starburst spectral features^{8,28}. The total stellar mass, estimated from fitting of the rest-frame ultraviolet-to-near-infrared spectral energy distribution, is $M_* = 5.5 \times 10^{10} M_\odot$, and analysis of the rest-frame near-ultraviolet/optical spectra show that 90% of the stellar emission is contributed by populations younger than 6 million years, representing 20% of the stellar mass. The rest-frame V-band size in the Hubble Space Telescope imaging was measured following methods presented elsewhere¹⁵, which we briefly describe here. To parameterize the compactness of the galaxy, we fitted Sérsic and Sérsic + PSF (point spread function) models to the galaxy, with the Sérsic index frozen to $n = 4$, and its nearby masked field (a $28'' \times 28''$ box centred on the galaxy) with GALFIT version 3 (ref. 29). The effective radius is $r_e = 94$ pc, with two-thirds of the rest-frame ultraviolet/optical light unresolved by the Hubble Space Telescope.

The bolometric luminosity was estimated through extrapolation of the mid-infrared (Wide-Field Infrared Survey Explorer 12 μm , 22 μm) photometry⁸, using appropriate spectral energy distribution templates that provide an estimate of the far-infrared emission, assuming these compact star-forming galaxies conform to the typical cool dust emission seen in star-forming galaxies³⁰. The SFR derived from the total (integrated over 8–1,000 μm) dust emission is in agreement with that derived from the fitting of the rest-frame 0.1–3 μm spectral energy distribution, with proper treatment of differential dust obscuration of the stellar emission^{8,31}. The measured $L_{\text{IR}} \approx 1.2 \times 10^{39}$ W corresponds to an SFR of $260 M_\odot \text{ yr}^{-1}$, assuming a Chabrier initial mass function³² (a Salpeter initial mass function would increase the SFR by 80%). If 50% of the star formation occurs within the effective radius measured above, then the projected SFR density is $\Sigma_{\text{SFR}} \approx 4,700 M_\odot \text{ yr}^{-1} \text{ kpc}^{-2}$. As a guide to the level of uncertainty in derived properties, which depend largely on template fitting, both the stellar mass and infrared luminosity are estimated to be accurate to within a factor of two.

IRAM observations and data reduction. SDSS J0905+57 was observed on 6–12 May 2013 and 10 December 2013 as part of IRAM Plateau de Bure Interferometer projects W09A and X09C. PdBI was in compact (D) configuration with baselines of 25–140 m. We used the WideX correlator, targeting the redshifted CO(2–1) line at $\nu_{\text{obs}} \approx 134$ GHz in the 2 mm band, recording dual polarization. The mean system temperature was $T_{\text{sys}} = 80$ –120 K and precipitable water vapour was in the range 2–6 mm. The sources 3C84, 3C279, 3C454.3 2200+420 and 0851+202 were used for bandpass calibration, and sources 0954+658 and 0917+624 were used for phase/amplitude calibration. We rejected scans for which the phase root mean square deviated more than 45° from the calibration solution. Finally, the source MWC349 was used for flux calibration (accuracy 5%–10% at 2 mm). The final root mean squared noise in 20 MHz (45 km s^{-1}) channels is $\sigma = 0.4$ mJy. The package GILDAS³³ was used for data calibration, mapping and analysis.

Line detection. We first mapped the u - v visibilities into the image plane to create a spectral cube from which we extract a spectrum from a single $0.6''$ pixel at the phase tracking centre. This spectrum is shown in Fig. 1, with the CO(2–1) emission line strongly detected at the expected frequency. The line is well modelled by a single Gaussian, with peak $S = 2.8 \pm 0.3$ mJy and $\sigma_{\text{FWHM}} = 200 \pm 35$ km s^{-1} . Uncertainties on the profile fitting parameters such as line centres and widths were estimated in the following way: first, we estimated the uncertainty per channel by extracting spectra from 100 random locations close to the phase tracking centre, but in line-free parts of the data cube, and then evaluated the root mean squared variation in the signal for each channel. Under the assumption that the noise is drawn from a Gaussian distribution \mathcal{N} with a mean of zero and a 1σ scale equivalent to the root mean square, we generated 1,000 realizations of the target spectrum, each time adding flux drawn randomly from \mathcal{N} to each channel. The integrated flux and line fits were re-evaluated for each of the 1,000 realizations, and the standard deviation of the derived values were taken to be the 1σ uncertainties on the fitting parameters. The integrated line flux over the 200–1,000 km s^{-1} wing is $\Delta V = 0.26 \pm 0.07$ Jy km s^{-1} , corresponding to $L'_{\text{CO}} = (1.8 \pm 0.5) \times 10^9$ K $\text{km s}^{-1} \text{ pc}^2$. When integrated over the core line, $|\Delta V| < 200$ km s^{-1} , the integrated line fluxes in the extended and unresolved components (see below) are $\Delta V = 0.46 \pm 0.20$ Jy km s^{-1} and $\Delta V = 0.56 \pm 0.12$ Jy km s^{-1} , respectively, corresponding to $L'_{\text{CO}} = (3.1 \pm 1.3) \times 10^9$ K $\text{km s}^{-1} \text{ pc}^2$ and $L'_{\text{CO}} = (3.8 \pm 0.8) \times 10^9$ K $\text{km s}^{-1} \text{ pc}^2$.

Red, spatially offset high-velocity wing emission. The spectrum shown in Fig. 1 reveals evidence of CO emission redward of the core Gaussian line in a broad wing that extends to approximately 1,000 km s^{-1} . Averaging the one-dimensional spectrum shown in Fig. 1 over 200–1,000 km s^{-1} yields an average flux density of $S = 0.34 \pm 0.09$ mJy for this feature. When the spectral cube is collapsed over the same channels, the peak of this wing emission is spatially offset $1.2 \pm 0.3''$ from the peak of the core line, with an average flux density of $S = 0.43 \pm 0.09$ mJy (Fig. 2). The positional uncertainty was estimated as $\sigma = 0.3 \times \text{FWHM}/\text{SNR}$, where SNR is the signal-to-noise ratio^{34,35}, which is consistent with the positional error derived from

a simulation involving the input of 1,000 model sources of the same flux level into the noise map. The scatter (standard deviation) in recovered positions is $0.3''$ in right ascension and declination with no systematic positional offset.

An alternative way of assessing the significance of this feature is to use the random noise realizations of the data cube as described above, but excluding the source; that is, just considering the noise component. This takes into account the possibility that the noise in consecutive channels might be correlated. We then evaluated the rate at which we measure flux densities of $S \geq 0.43$ mJy averaged over the same channels as the observed wing feature. Using 10^7 noise realizations, we found a rate of 1.4×10^{-6} , consistent with the 4.8σ significance of the detection (we found the same rate of negative fluctuations of equivalent magnitude).

Resolved core line emission. When averaged over the core Gaussian line, the map suggests that the CO emission is extended compared to the beam (Fig. 2). To verify this, and to evaluate the size of the emitting region, we examined the velocity averaged signal amplitude as a function of baseline separation (that is, synthetic aperture size) in the u - v plane, since an unresolved source will have a flat amplitude–radius profile. We averaged over the full linewidth, $|\Delta V| < \text{FWHM}$, approximately corresponding to the FWZI (400 km s^{-1}). Extended Data Fig. 1 shows that the flux distribution, evaluated as the average amplitude in radial bins of width 50 m, deviates from a flat distribution at baseline separations shorter than approximately 100 m. We see evidence for this in both independent observations of the target. As described above, we imposed strict flagging throughout, rejecting scans for which the phase root mean square deviated more than 45° from the calibration solution, and therefore any smearing of the signal related to phase calibration errors will occur on scales of approximately $1.5''$ (roughly half the size of the synthesized beam). This is a strong indication that the extended emission is real and not a result of seeing; however, as an additional check, we examined the profiles of the phase calibrators themselves.

If the extended emission is due to a phase calibration issue, in which point source emission is artificially smeared out, then we would expect to see a similar extended profile around the calibrators as well as the target. The first test we performed was to measure the average amplitude as a function of u - v separation for the main calibrator common to both projects: source 0917+624. Extended Data Figs 2 and 3 summarize the results with the map and amplitude–radius profile of the main calibrator, with the master phase calibration solution applied, indicating consistency with an unresolved source. During project X09C, two calibrators were observed: 0917+624 and 0954+658. This gave us the opportunity to derive a phase calibration solution excluding one of these sources (0954+658) and then apply that solution to the excluded source. This is a robust test, since 0954+658 has not contributed to the phase solution, and can be treated as an independent source. We show the map for 0917+624 and 0954+658 (with the latter ‘blindly’ treated with the phase solution derived from the main calibrator) in Extended Data Fig. 2, and the corresponding amplitude–radius profiles in Extended Data Fig. 3. Again, the results indicate unresolved sources, which would not be the case if phase calibration errors are responsible for the extended emission observed in the target source. These results give us further confidence that the extended emission we observe in SDSS J0905+57 is real.

It is common practice to model partially resolved emission with a Gaussian profile, the half-power size of which can be used to characterize the size of the emitting region. We fitted the flux distribution with a combination of a Gaussian profile and a point source:

$$S(r) = S_0 \exp\left(-\frac{(\pi r b)^2}{\ln(2)}\right) + S_1 \quad (1)$$

where $r = (u^2 + v^2)^{1/2}$, S_0 is the peak amplitude of the extended component, b is the half-width at half-power size and S_1 is the amplitude of the unresolved component. To improve the signal-to-noise ratio, we had to average the visibilities into three bins, so to reduce the number of free parameters in the model, we made the following assumptions: first, we assumed that the flux density on the longest baselines is representative of the flux density of the unresolved component, $S_1 = 1.6 \pm 0.3$ mJy. We then fixed $S_0 = (2.8 - S_1)$ mJy, corresponding to the peak emission measured in the spectrum described above, corrected for a contribution from an unresolved component. With these parameters fixed, we then performed a χ^2 minimization between the model and the data to find the best-fitting half-power size, b . To find the range of acceptable values for the fit, taking into account the uncertainties on the assumptions for S_0 and S_1 , we performed a Monte Carlo simulation, re-fitting the data by sampling S_0 and S_1 from Gaussian distributions with widths concordant with the 1σ errors on the flux densities. We performed 1,000 trials and took the mean and standard deviation in best-fit for each trial b as the final size estimate. We found $b = 62 \pm 21$ m or $\theta = 1.6^{+0.8}_{-0.4}$ arcseconds for $\lambda = 2.23$ mm, with a best-fit $\chi^2 = 0.07$ for one degree of freedom, implying that the data are over-fitted. Additional observations that would allow us to increase the number of bins in the amplitude–radius plane would improve our constraints on the size of the extended emission. Nevertheless, the χ^2 for the null hypothesis (that the source is unresolved) is $\chi^2 = 20.8$,

with the $\Delta\chi^2$ corresponding to a significance of 4.7σ . Thus, we can rule out the point-source-only model with reasonable confidence.

Feedback energetics. SDSS J0905+57 is forming stars close to the theoretical upper limit for radiation pressure (Eddington-limited) star formation, with $L_{\text{IR}} \approx L_{\text{Edd}} \propto GcL_{\text{CO}}/\kappa$ for optically thick dust emission ($\tau_{100\mu\text{m}} > 1$). Here L_{IR} is the measured total infrared luminosity (see above), L_{Edd} is the Eddington luminosity, c is the speed of light in vacuum, G is the gravitational constant and κ is the Rosseland-mean dust opacity ($\kappa \approx 1,000f_{\text{dg}}\text{cm}^2\text{g}^{-1}$, where f_{dg} is the dust-to-gas mass ratio, typically $1/50 - 1/150$; ref. 10). The high SFR density of SDSS J0905+57 is above the threshold necessary for launching a radiation-pressure-driven wind^{8–10,36}.

The clear evidence that the galaxy has a high velocity outflow, traced by Ca II, Fe II and Mg II (A.M.D.-S., J.M., C.A.T., A.L.C., R.C.H., G.H.R., P.H.S. & J.E.G., manuscript in preparation) is an unambiguous signature of an energetic, gaseous outflow launched in the recent past. The observations presented here imply that molecular gas has also been ejected by feedback processes. In the following we estimate the rate of momentum input from the starburst and compare this to the energetics of the wind in order to determine whether radiation pressure is a viable power source for the outflow. Considering the gas associated with the high-velocity wing, we determine the mass outflow rate as follows:

$$\dot{M}_{\text{H}_2} = \frac{M_{\text{H}_2} v}{r} \quad (2)$$

where M_{H_2} is the gas mass in the outflow, v is the outflow velocity, and r is the radius of the outflow. We make the conservative assumption that the gas is travelling at $1,000\text{ km s}^{-1}$, assuming that the maximum velocity extent of the wing is representative of the deprojected (that is, three-dimensional) velocity of the outflow²². The mass outflow rate is $\dot{M}_{\text{H}_2} = 80 \pm 25 M_{\odot} \text{ yr}^{-1}$ and it follows that the kinetic power in the outflow is

$$P_K = \frac{\dot{M}_{\text{H}_2} v^2}{2} \quad (3)$$

which yields $P_K = (2.6 \pm 0.8) \times 10^{36} \text{ W}$, several orders of magnitude lower than the bolometric luminosity. Again, it is important to highlight that the error bars do not reflect the systematic uncertainty from the choice of α that we use to estimate M_{H_2} from L'_{CO} . In these calculations we have assumed an $\alpha = 0.34 M_{\odot} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$ that could be appropriate for the optically thin conditions in a turbulent molecular outflow¹⁷. The conversion factor z is defined as $M_{\text{H}_2}/L'_{\text{CO}}$ for $J = 0$, so a correction is required for higher-order transitions (as are often measured in high-redshift galaxies) to account for the shape of the spectral line energy distribution that describes the excitation of the gas. We have no constraints on the excitation state of the molecular gas in this galaxy (a wider range of transitions will be needed to constrain the spectral line energy distribution), and so for all components we have assumed thermalized CO emission such that $r_{21} = L'_{\text{CO}(2-1)}/L'_{\text{CO}(1-0)} = 1$.

We argue that the outflow can be driven by radiation pressure from the compact starburst, so it is critical to assess the momentum injection available to drive cold gas. In the single scattering limit a starburst's radiation pressure scales with the bolometric luminosity, L/c (refs 23 and 24). For SDSS J0905+57 we find a momentum rate $\dot{p}_{\text{rad}} \approx 4 \times 10^{30} \text{ N}$. The momentum flux in the wind is $v\dot{M}_{\text{H}_2} \approx (4.8 \pm 1.9) \times 10^{30} \text{ N}$, implying that the starburst could be driving the outflow through stellar radiation pressure alone, even in the single scattering limit. Given that we have neglected other driving forces such as supernovae ram pressure, which could provide an additional factor of about three times the momentum input from radiation pressure²⁴, we conclude that it is plausible that a central compact starburst of this magnitude could drive the molecular outflow we observe.

How does the energy of the molecular outflow compare to the high-velocity outflow of warmer ionized gas traced by Mg II? There are several uncertainties in deriving an energy using the Mg II outflow, the most serious of which is the lack of any useful constraints on the geometry and scale of the outflow (although the saturated Mg II lines suggest a high covering factor that is possibly consistent with a shell geometry). Other uncertainties include the use of saturated Mg II lines, uncertain Mg^{+}/Mg ionization and dust depletion corrections and Mg/H abundance. Nevertheless, for a shell of radius r the mass scales as

$$M \approx 5.5 \times 10^8 M_{\odot} \left(\frac{r}{5 \text{ kpc}} \right)^2 \left(\frac{N_{\text{H}}}{2 \times 10^{20} \text{ cm}^{-2}} \right)$$

and the kinetic energy as

$$E \approx 3.4 \times 10^{51} \text{ J} \left(\frac{r}{5 \text{ kpc}} \right)^2 \left(\frac{N_{\text{H}}}{2 \times 10^{20} \text{ cm}^{-2}} \right) \left(\frac{v}{2500 \text{ km s}^{-1}} \right)^2$$

Thus, for a shell of radius 5 kpc and column of $N_{\text{H}} = 2 \times 10^{20} \text{ cm}^{-2}$, there is comparable mass and approximately six times more energy in the ionized outflow compared

to the molecular outflow, indicating that a contribution from supernovae is required to drive the ionized wind. A possible scenario is that the blow-out phase of the final-stage merger involves two key stages: radiative feedback that first drives cold gas out of the bulge and into the halo, clearing low-density 'escape paths' for warmer gas that is then driven out by supernova detonations, several million years later⁹.

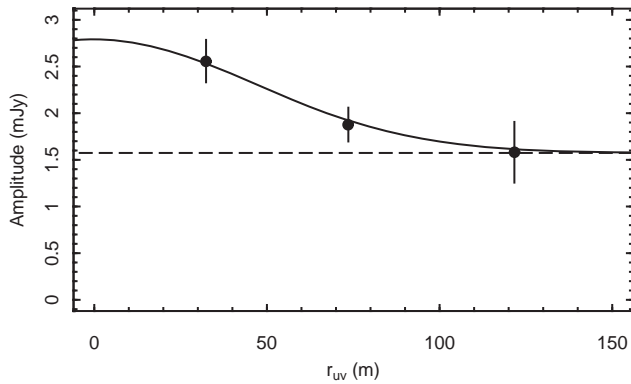
What of the other extended, low-velocity CO component associated with the core line? It is difficult to infer feedback energetics associated with this gas, which is extended on scales of 12 kpc. This gas is moving with lower velocity than the outflow (that is, similar to the circular velocity, $(GM/R)^{0.5} \approx 150 \text{ km s}^{-1}$ with $M(r < R) = 6 \times 10^{10} M_{\odot}$ dominated by baryons and $R = 12 \text{ kpc}$). If this gas were ejected in a shell that has been radiatively driven into the halo in a previous outflow event, it is likely to have fragmented and decelerated; at late times the cold clouds are subject to lower radiation and ram-pressure⁷. In this case an instantaneous measure of mass outflow rate and its connotations loses the intended meaning, and the previously expelled cold gas merely represents the time-integrated feedback history.

Arguments for starburst feedback and against AGN feedback. The arguments for star formation as opposed to AGN feedback for this galaxy and others like it are discussed in previous work^{8,15}. For this source, we have shown that the compact starburst can produce the observed outflows, but cannot conclusively rule out that they were launched by an AGN that has since 'switched off'. However, the larger population of compact starbursts shows ubiquitous outflows with no correlation with AGN activity, pointing to star formation as the most likely driver for the observed feedback in this and similar systems. Furthermore, there are two main pieces of observational evidence that imply that SDSS J0905+57 does not contain an energetically dominant AGN:

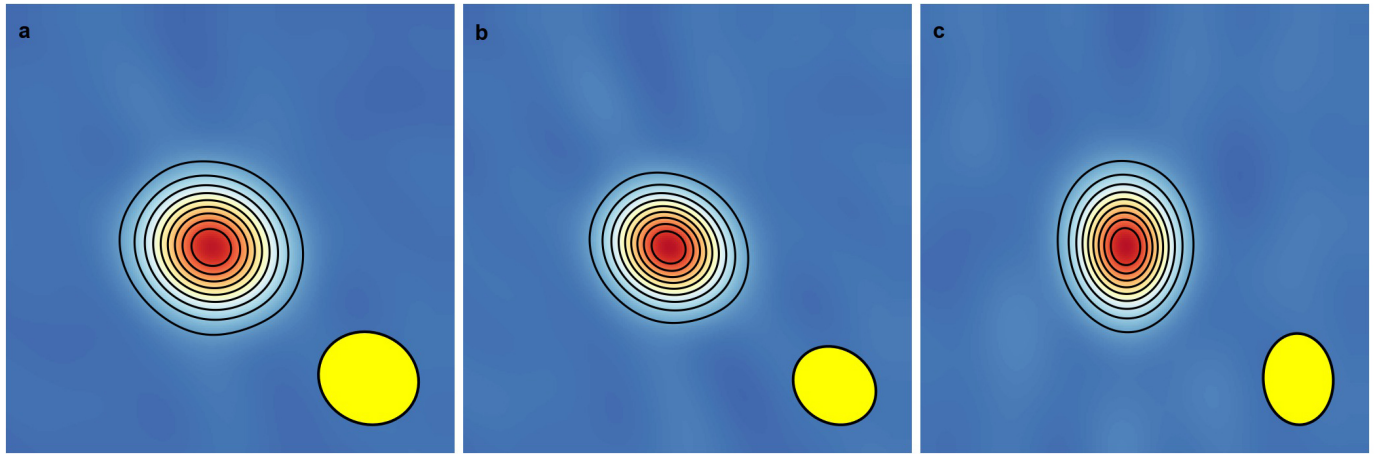
(1) AGN activity is usually diagnosed via high-excitation emission lines, especially when X-ray observations are unavailable (as is the case for SDSS J0905+57) because it is typically assumed that the AGN is the only possible source for considerable numbers of high-energy photons. We have observational constraints on the emission lines $[\text{O III}] \lambda = 5,007 \text{ \AA}$, $[\text{Ne III}] \lambda = 3,869 \text{ \AA}$ and $[\text{Ne V}] \lambda = 3,426 \text{ \AA}$. We measure $\log([[\text{O III}]/\text{H}\beta]) = 0.19$ and $L_{[\text{O III}]} = 7 \times 10^{34} \text{ W}$. No $[\text{Ne V}] \lambda = 3,426 \text{ \AA}$ is detected ($L_{[\text{Ne V}]} < 2.25 \times 10^{33} \text{ W}$, 1σ upper limit), and the $[\text{Ne III}]$ line is weak, $L_{[\text{Ne III}]} = 1.64 \times 10^{33} \text{ W}$ (3σ detection). Very compact starburst galaxies can produce slightly elevated excitation emission lines with these observed properties³⁷ and these observations are consistent with a starburst-dominated system at $z \approx 0.7$.

(2) Mid-infrared observations can be used to assess obscured AGN activity. The $3.6\text{--}4.5 \mu\text{m}$ colour from Spitzer observations of SDSS J0905+57 is $[3.6] - [4.5] = 0.48$ (in Vega magnitudes), which is in a transition region between star-forming galaxies and AGNs^{38,39}. Note that the $5.8\text{--}8.0 \mu\text{m}$ colour was unavailable in the Spitzer Warm Mission observations. We can also estimate L_{bol} from $L_{[\text{O III}]}$, in the limit where all the $[\text{O III}]$ is produced from an AGN. Keeping in mind that $L_{[\text{O III}]}$ is probably heavily contaminated by star formation, using $L_{\text{bol}}/L_{[\text{O III}]} = 600$ (ref. 40), we find $L_{\text{bol}} = 4 \times 10^{37} \text{ W}$, which is approximately 4% of the total infrared luminosity.

28. Tremonti, C. A., Moustakas, J. & Diamond-Stanic, A. M. The discovery of 1000 km s^{-1} outflows in massive poststarburst galaxies at $z=0.6$. *Astron. J.* **663**, L77–L80 (2007).
29. Peng, C. Y., Ho, L. C., Impey, C. D. & Rix, H.-W. Detailed structural decomposition of galaxy images. *Astron. J.* **124**, 266–293 (2002).
30. Chary, R. & Elbaz, D. Interpreting the cosmic infrared background: constraints on the evolution of the dust enshrouded star formation rate. *Astrophys. J.* **556**, 562–581 (2001).
31. Moustakas, J. et al. PRIMUS: constraints on star formation quenching and galaxy merging, and the evolution of the stellar mass function from $z = 0\text{--}1$. *Astrophys. J.* **767**, 50–84 (2013).
32. Kennicutt, R. C. & Evans, N. J. Star formation in the Milky Way and nearby galaxies. *Annu. Rev. Astron. Astrophys.* **50**, 531–608 (2012).
33. Guilloteau, S. & Lucas, R. Imaging at radio through submillimeter wavelengths. In *ASP Conf. Proc.* (eds Mangum, J. G. & Radford, S. J. E.) Vol. 217 (Astronomical Society of the Pacific, 2000).
34. Condon, J. J. Errors in elliptical gaussian FITS. *Publ. Astron. Soc. Pacif.* **109**, 166–172 (1997).
35. Ivison, R. J. et al. The SCUBA Half Degree Extragalactic Survey—III. Identification of radio and mid-infrared counterparts to submillimetre galaxies. *Mon. Not. R. Astron. Soc.* **380**, 199–228 (2007).
36. Geach, J. E. et al. A redline starburst: CO(2-1) observations of an Eddington-limited galaxy reveal star formation at its most extreme. *Astrophys. J. Lett.* **767**, 17–22 (2013).
37. Kewley, L. J. et al. Theoretical evolution of optical strong lines across cosmic time. *Astrophys. J.* **774**, 100–117 (2013).
38. Donley, J. L. et al. Identifying luminous active galactic nuclei in deep surveys: revised IRAC selection criteria. *Astrophys. J.* **748**, 142–164 (2012).
39. Stern, D. et al. Mid-infrared selection of active galactic nuclei with the Wide-Field Infrared Survey Explorer. I. Characterizing WISE-selected active galactic nuclei in COSMOS. *Astrophys. J.* **753**, 30–48 (2012).
40. Kauffmann, G. & Heckman, T. M. Feast and famine: regulation of black hole growth in low-redshift galaxies. *Mon. Not. R. Astron. Soc.* **397**, 135–147 (2009).

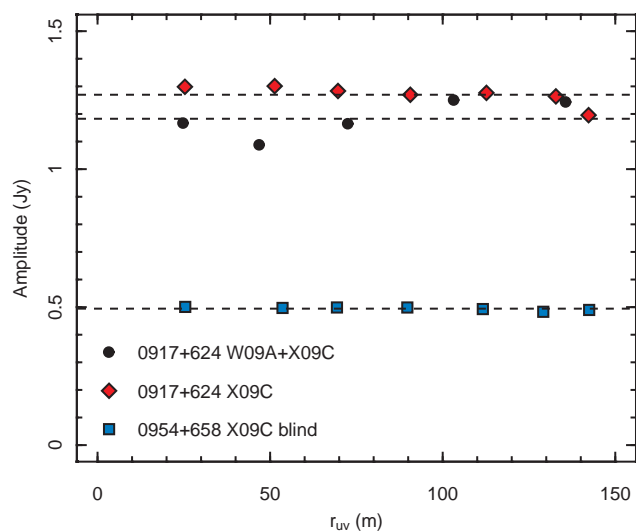


Extended Data Figure 1 | Average signal amplitude as a function of baseline separation for CO emission over the core line, $\Delta V = \pm 200 \text{ km s}^{-1}$. This reveals a significant deviation from a flat profile, indicating that the CO emission is partially resolved. The profile is accurately modelled by a combination of a point source (a delta function in the u - v plane) and a Gaussian profile with half-power radius of $\sim 2''$. A point-source-only model can be ruled out at the 4.7σ level. Error bars show the 1σ confidence range, and are derived as $w^{-0.5}$, where w is the weight $w = \Delta v / T_{\text{sys}}^{\text{ab}}$, where Δv is the channel width, t is the integration time and $T_{\text{sys}}^{\text{ab}}$ is the product of the system temperature of two antennas. r_{uv} is the separation of antennas in the u - v plane.



Extended Data Figure 2 | Clean maps of the phase calibrators 0917+624 and 0954+658. The FWHM beam shape is indicated as a yellow ellipse. Contours are at levels of 10% to 90% of the peak flux, with the 90% flux line closest to the centre. **a**, Phases were calibrated with a solution derived from calibrators observed over both projects. **b**, The same phase solution was applied

to observations of 0917+624 observed in project X09C only. **c**, The calibrator 0954+658 was calibrated with a phase solution derived from 0917+624 observed during project X09C only. All sources have a profile that is matched to the synthesized beam, with no evidence of extended emission.



Extended Data Figure 3 | Circularly averaged amplitude-radius profiles of the maps shown in Extended Data Fig. 2 in the u - v plane. All profiles are consistent with unresolved emission (flat profiles). Note that observing conditions were slightly better during project X09C than during W09A. Dashed lines indicate the mean amplitude. 0954+658 X09C ‘blind’ refers to the calibration solution derived from source 0917+624 only (see Methods).

Nonlinear lattice dynamics as a basis for enhanced superconductivity in $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$

R. Mankowsky^{1,2,3*}, A. Subedi^{4*}, M. Först^{1,3}, S. O. Mariager⁵, M. Chollet⁶, H. T. Lemke⁶, J. S. Robinson⁶, J. M. Glowinski⁶, M. P. Minitti⁶, A. Frano⁷, M. Fechner⁸, N. A. Spaldin⁸, T. Loew⁷, B. Keimer⁷, A. Georges^{4,9,10} & A. Cavalleri^{1,2,3,11}

Terahertz-frequency optical pulses can resonantly drive selected vibrational modes in solids and deform their crystal structures^{1–3}. In complex oxides, this method has been used to melt electronic order^{4–6}, drive insulator-to-metal transitions⁷ and induce superconductivity⁸. Strikingly, coherent interlayer transport strongly reminiscent of superconductivity can be transiently induced up to room temperature (300 kelvin) in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ (refs 9, 10). Here we report the crystal structure of this exotic non-equilibrium state, determined by femtosecond X-ray diffraction and *ab initio* density functional theory calculations. We find that nonlinear lattice excitation in normal-state $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ at above the transition temperature of 52 kelvin causes a simultaneous increase and decrease in the Cu–O₂ intra-bilayer and, respectively, inter-bilayer distances, accompanied by anisotropic changes in the in-plane O–Cu–O bond buckling. Density functional theory calculations indicate that these motions cause drastic changes in the electronic structure. Among these, the enhancement in the $d_{x^2-y^2}$ character of the in-plane electronic structure is likely to favour superconductivity.

The response of a crystal lattice to strong, resonant excitation of an infrared-active phonon mode can be described by separating the crystal Hamiltonian into its linear and nonlinear terms: $H = H_{\text{lin}} + H_{\text{NL}}$. The linear term $H_{\text{lin}} = \omega_{\text{IR}}^2 Q_{\text{IR}}^2/2$ describes harmonic oscillations about the equilibrium atomic positions, with ω_{IR} denoting the frequency and Q_{IR} the normal coordinate of the infrared-active mode. In the limit of lowest-order (cubic) coupling to other modes with generic coordinate Q_{R} , the nonlinear term can be written as $H_{\text{NL}} = \omega_{\text{R}}^2 Q_{\text{R}}^2/2 - a_{12} Q_{\text{IR}} Q_{\text{R}}^2 - a_{21} Q_{\text{IR}}^2 Q_{\text{R}}$. In this expression, a_{12} and a_{21} are anharmonic coupling constants. (See Methods section and Extended Data Figs 1 and 2 for details on quartic

coupling.) For a centrosymmetric crystal like $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$, $a_{12} Q_{\text{IR}} Q_{\text{R}}^2$ is zero because Q_{IR} is odd in symmetry whereas Q_{R}^2 is even and their product therefore vanishes.

Thus, the total Hamiltonian reduces to $H = \omega_{\text{IR}}^2 Q_{\text{IR}}^2/2 + \omega_{\text{R}}^2 Q_{\text{R}}^2/2 - a_{21} Q_{\text{IR}}^2 Q_{\text{R}}$, which results in a shift in the potential energy minimum along Q_{R} for any finite distortion Q_{IR}^* (Fig. 1a). Correspondingly, for a periodically driven Q_{IR} mode, the dynamics are described by the coupled equations of motion

$$\ddot{Q}_{\text{IR}} + 2\gamma_{\text{IR}} \dot{Q}_{\text{IR}} + \omega_{\text{IR}}^2 Q_{\text{IR}} = f(t) + 2a_{21} Q_{\text{IR}} Q_{\text{R}}$$

and

$$\ddot{Q}_{\text{R}} + 2\gamma_{\text{R}} \dot{Q}_{\text{R}} + \omega_{\text{R}}^2 Q_{\text{R}} = a_{21} Q_{\text{IR}}^2.$$

Figure 1b pictorially represents these dynamics. On resonant mid-infrared excitation of Q_{IR} , a unidirectional force is exerted along the normal coordinate Q_{R} , which is displaced by a magnitude proportional to Q_{IR}^2 . This effect remains sizeable only as long as Q_{IR} oscillates coherently, which occurs typically for several picoseconds.

We next discuss the specific case of $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$, which crystallizes in a centrosymmetric orthorhombic unit cell that has D_{2h} symmetry (point group), comprising bilayers of conducting CuO_2 planes separated by an insulating layer containing Y atoms and Cu–O chains that control the hole doping of the planes (Fig. 2a). The $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$ sample contained both O-rich and O-deficient chains, and exhibited short-range

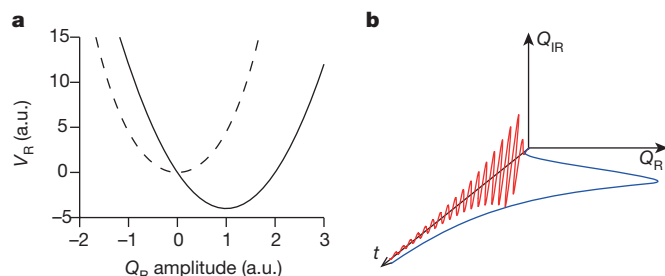


Figure 1 | Coherent nonlinear lattice dynamics in the limit of cubic coupling. **a**, A static distortion Q_{IR}^* shifts the equilibrium potential (V_{R} ; dashed line) of all modes Q_{R} that are coupled through $Q_{\text{IR}}^2 Q_{\text{R}}$ coupling, displacing the equilibrium position towards a new minimum (solid line). a.u., arbitrary units. **b**, The dynamical response of the two modes involves an oscillatory motion of the infrared mode (red line) and a directional displacement of Q_{R} (blue line). The displacement is proportional to Q_{IR}^2 and survives as long as Q_{IR} is coherent.

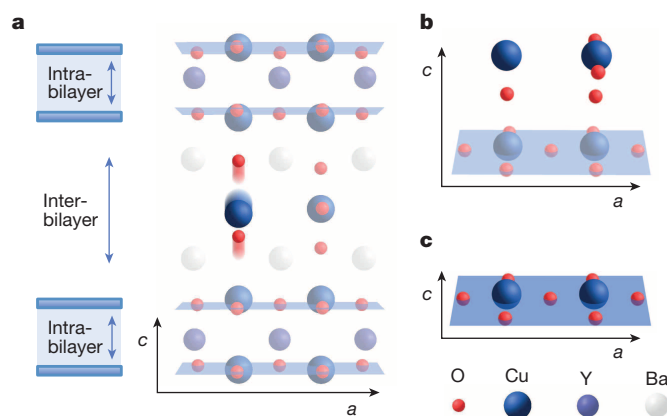


Figure 2 | Structure of $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$. **a**, Structure of orthorhombic $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$ and motions of the optically excited B_{1u} mode. The sketch on the left shows the two tunnelling regions respectively within and between the bilayers. **b**, Cu–O chains, which are either filled (Cu on right) or empty (Cu on left) in the ortho-II structure. **c**, Superconducting CuO_2 planes (blue).

¹Max Planck Institute for the Structure and Dynamics of Matter, 22761 Hamburg, Germany. ²University of Hamburg, 22761 Hamburg, Germany. ³Center for Free-Electron Laser Science (CFEL), 22761 Hamburg, Germany. ⁴Centre de Physique Théorique, École Polytechnique, CNRS, 91128 Palaiseau Cedex, France. ⁵Swiss Light Source, Paul Scherrer Institut, 5232 Villigen, Switzerland. ⁶Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park 94025, California, USA. ⁷Max Planck Institute for Solid State Research, 70569 Stuttgart, Germany. ⁸Materials Theory, Eidgenössische Technische Hochschule Zürich, 8093 Zürich, Switzerland. ⁹Collège de France, 11 place Marcelin Berthelot, 75005 Paris, France. ¹⁰Département de Physique de la Matière Condensée (MaNEP), Université de Genève, 1211 Genève, Switzerland. ¹¹Department of Physics, University of Oxford, Clarendon Laboratory, Oxford OX1 3PU, UK.

*These authors contributed equally to this work.

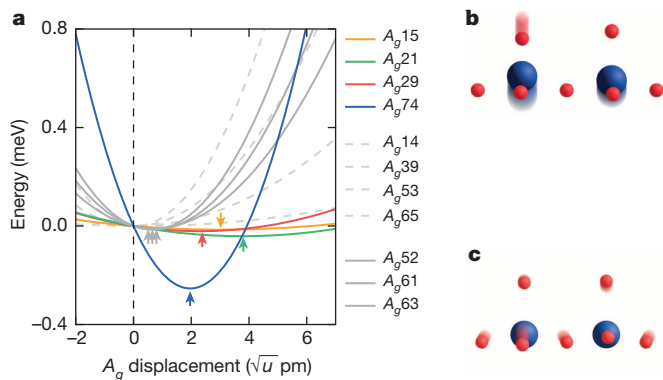


Figure 3 | First-principles calculations of cubic coupling between 11 A_g modes and the driven B_{1u} mode. **a**, Energy potentials of all A_g modes for a frozen B_{1u} displacement of 0.14 Å/u (u , atomic mass unit), corresponding to a change in apical O–Cu distance of 2.2 pm. The x axis is the amplitude of the A_g eigenvector. Arrows indicate the potential minima. **b**, There is strong coupling to the $A_g(15, 21, 29, 74)$ modes, which involves a decrease in the apical O–Cu distance and an increase in in-plane buckling. **c**, The $A_g(52, 61, 63)$ modes are weakly coupled and govern a breathing motion of the oxygen atoms in the CuO_2 plane.

ortho-II ordering of the vacancies (Fig. 2b). We note also that the in-plane O–Cu–O bonds are buckled (Fig. 2c).

In our experiments, mid-infrared pump pulses of ~ 300 fs duration were focused to a maximum fluence of $\sim 4 \text{ mJ cm}^{-2}$ and a peak electric field of $\sim 3 \text{ MV cm}^{-1}$. These pulses were polarized along the c axis of $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$ and tuned to resonance with the same 670 cm^{-1} frequency ($\sim 15 \text{ μm}$, 83 meV) B_{1u} infrared-active mode¹¹ (Fig. 2a) that was previously shown by means of time-resolved terahertz spectroscopy to enhance interlayer superconducting coupling^{9,10}.

In analysing the nonlinear lattice dynamics caused by this excitation, we note that the nonlinear term $a_{21}Q_{B_{1u}}^2Q_R$, where $Q_{B_{1u}}$ denotes the normal coordinate of the infrared-active mode, is non-zero only if Q_R is of A_g symmetry, because the square of the irreducible representation of B_{1u} is A_g . Thus, only A_g modes can couple to the optically driven B_{1u}

motion shown in Fig. 2a. $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$ has 72 optical phonon modes, of which 33 are Raman active. These can be further divided into 22 B_g modes, which break in-plane symmetry, and 11 A_g modes, which preserve the symmetry of the unit cell (Extended Data Fig. 3). The geometries of these 11 A_g modes and their respective coupling strengths to the driven B_{1u} mode were computed using first-principles density functional theory (DFT) calculations within the local-density approximation. At the 3 MV cm^{-1} field strength of the mid-infrared pump pulse, we expect a peak amplitude for the B_{1u} motion corresponding to a 2.2 pm increase in the distance between the apical O atom and the chain Cu atom, which was used as a basis to calculate energy potentials of the A_g modes for a frozen distortion of this magnitude (Fig. 3a). Only four phonon modes $A_g(15, 21, 29, 74)$ were found to couple strongly to the driven B_{1u} mode, all involving a concerted distortion of the apical O atoms towards the CuO_2 plane and an increase in Cu–O buckling (Fig. 3b). (The numbers in parentheses denote the indices of the phonon modes, sorted in order of increasing frequency.) The calculations also predict weak coupling to three further modes $A_g(52, 53, 61)$, consisting of breathing motion of the O atoms in the plane (Fig. 3c). The remaining four modes $A_g(14, 39, 53, 65)$ do not couple to the B_{1u} mode (Extended Data Table 1).

To experimentally determine the absolute amplitude of these distortions under the conditions relevant for enhanced superconductivity^{9,10}, we measured time-resolved X-ray diffraction using 50 fs, 6.7 keV pulses from the Linac Coherent Light Source free-electron laser, which was synchronized with the optical laser that generated the mid-infrared pump pulses. Changes in diffraction intensity were recorded for four Bragg peaks at a base temperature of 100 K, which is above the equilibrium transition temperature $T_c = 52 \text{ K}$. These peaks were observed to either increase or decrease promptly after excitation (Fig. 4) and to relax within the same timescale as the changes in the terahertz optical properties^{9,10}. For each Bragg reflection, we calculated changes in diffraction as functions of $Q_{B_{1u}}$ amplitude considering a displacement of only the four dominant Raman modes $A_g(15, 21, 29, 74)$ or all 11 modes in Fig. 3, taking into account the relative coupling strengths. We simultaneously fitted the four experimental diffraction curves using only two free parameters: the amplitude of the directly driven B_{1u} motion and the relative contributions of two exponential relaxation components

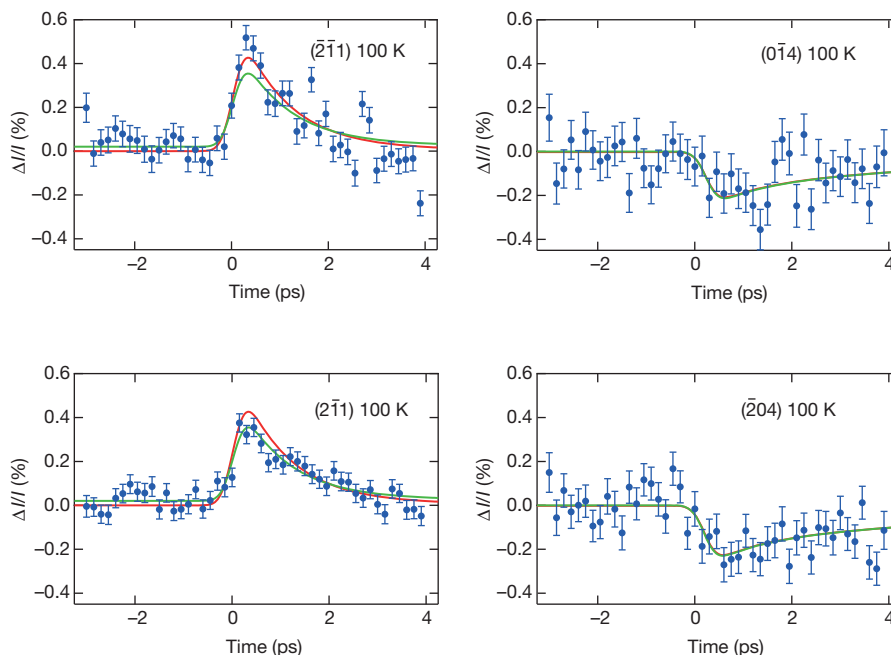


Figure 4 | Time-dependent diffracted peak intensity (I) for four Bragg reflections. A displacive lattice distortion is observed. The experimental data are fitted (solid curves) by adjusting the B_{1u} amplitude and the relative strength of the two relaxation channels ($\tau_1 = 1 \text{ ps}$, $\tau_2 = 7 \text{ ps}$) extracted from the optical

experiments of refs 9, 10. The relative amplitudes and signs of the curves are determined from the calculated structure using only the four most strongly coupled modes (green) or all A_g modes (red). Error bars, 1σ (67% confidence interval).

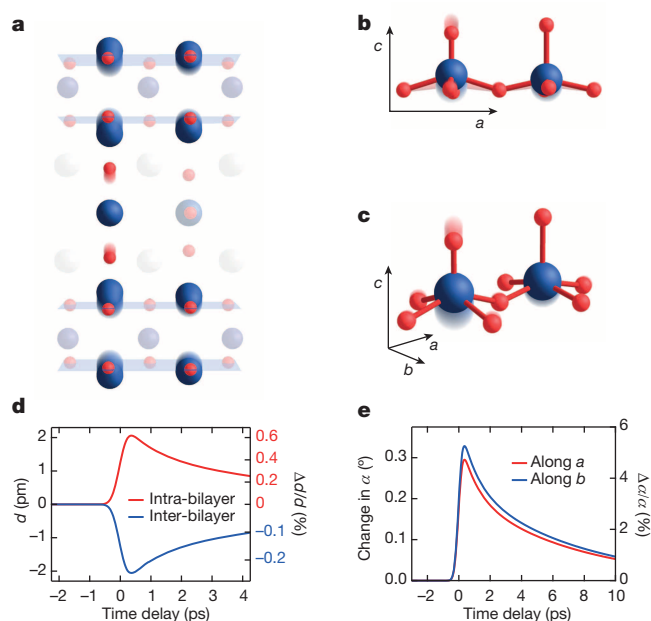


Figure 5 | Transient lattice structure. **a–c**, We find a concerted displacive lattice distortion (**b, c**) with a decrease in the apical O–Cu distances by 2.4 pm at O-deficient sites and an increase in O–Cu–O buckling. **d**, The intra-bilayer distance increases and the inter-bilayer distance decreases. Here the copper atoms of the planes at O-deficient chain sites (at left in **a**) are used to define the positions of the planes. **e**, The in-plane buckling angle α of the O–Cu–O bond increases by 5% along both a and b at oxygen-deficient sites.

($\tau_1 = 1$ ps, $\tau_2 = 7$ ps) extracted from the terahertz measurements^{9,10}. Very similar results were found when considering only the four dominant modes or all modes (Fig. 4, green and red fitting curves).

The transient lattice structure determined from these fits involves the following elements. First, we observe a decrease in the distance between the apical O and the Cu atoms of the superconducting planes (Fig. 5). This motion is far smaller than, and is opposite in sign to, the difference in the static apical O positions between La- and Hg-based copper oxide superconductors, for which T_c is greater at equilibrium¹². Therefore, the transient enhancement of superconducting transport cannot be explained by this analogy. More suggestively, the Cu atoms are simultaneously driven away from one another within the bilayers and towards one another between different bilayers. This spatially staggered motion is approximately 0.63% of the equilibrium intra-bilayer distance (Fig. 5a) and qualitatively follows the decrease in intra-bilayer tunnelling and the increase in inter-bilayer tunnelling^{9,10}. Finally, an anisotropic 0.32° increase in the in-plane O–Cu–O buckling (different along the a and b axes) is observed (Extended Data Table 2).

Although the Josephson coupling in layered copper oxide superconductors involves many microscopic parameters that are not taken into account here^{13–15}, DFT calculations in the distorted crystal structure were used to assess the salient effects on the electronic properties (Extended Data Figs 4–6). Our calculations predict a decrease of a few tens of millielectronvolts in the energy of the O-deficient chain bands. Because at equilibrium these bands are very close to the Fermi energy level, this small shift strongly reduces the hybridization of the chains with the plane Cu orbital, leading to a DFT Fermi surface with a stronger Cu $d_{x^2-y^2}$ character and higher hole doping. This effect is likely to favour superconductivity. We also speculate that as the DFT Fermi surface changes shape and size, it is quite possible that the charge-density wave order may also be destabilized^{16–18}, which would also aid superconductivity. The present calculations will serve as a starting point for a full many-body

treatment, to be complemented by more exhaustive experimental characterizations of the transient electronic structure.

More generally, we see nonlinear phononics as a new tool for dynamical materials discovery, with optical lattice control providing a perturbation—analogue to strain, fields or pressure—that can induce exotic collective electronic behaviour. Knowledge of the non-equilibrium atomic structure from ultrafast X-ray crystallography, which we provide here, is the essential next step towards engineering such induced behaviour at equilibrium.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 April; accepted 19 September 2014.

1. Först, M. *et al.* Nonlinear phononics as an ultrafast route to lattice control. *Nature Phys.* **7**, 854–856 (2011).
2. Först, M. *et al.* Displacive lattice excitation through nonlinear phononics viewed by femtosecond X-ray diffraction. *Solid State Commun.* **169**, 24–27 (2013).
3. Subedi, A., Cavalleri, A. & Georges, A. Theory of nonlinear phononics for coherent light control of solids. *Phys. Rev. B* **89**, 220301 (2014).
4. Tobey, R. I. *et al.* Ultrafast electronic phase transition in $\text{La}_{1/2}\text{Sr}_{3/2}\text{MnO}_4$ by coherent vibrational excitation: evidence for nonthermal melting of orbital order. *Phys. Rev. Lett.* **101**, 197404 (2008).
5. Först, M. *et al.* Driving magnetic order in a manganite by ultrafast lattice excitation. *Phys. Rev. B* **84**, 241104 (2011).
6. Först, M. *et al.* Melting of charge stripes in vibrationally driven $\text{La}_{1.875}\text{Ba}_{0.125}\text{CuO}_4$: assessing the respective roles of electronic and lattice order in frustrated superconductors. *Phys. Rev. Lett.* **112**, 157002 (2014).
7. Rini, M. *et al.* Control of the electronic phase of a manganite by mode-selective vibrational excitation. *Nature* **449**, 72–74 (2007).
8. Faust, D. *et al.* Light-induced superconductivity in a stripe-ordered cuprate. *Science* **331**, 189–191 (2011).
9. Kaiser, S. *et al.* Optically induced coherent transport far above T_c in underdoped $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. *Phys. Rev. B* **89**, 184516 (2014).
10. Hu, W. *et al.* Optically enhanced coherent transport in $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$ by ultrafast redistribution of interlayer coupling. *Nature Mater.* **13**, 705–711 (2014).
11. Homes, C. C. *et al.* Optical properties along the c -axis of $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ for $x = 0.50 \rightarrow 0.95$ evolution of the pseudogap. *Physica C* **254**, 265–280 (1995).
12. Pavarini, E. *et al.* Band-structure trend in hole-doped cuprates and correlation with $T_{c\text{max}}$. *Phys. Rev. Lett.* **87**, 047003 (2001).
13. Shah, N. & Millis, A. J. Superconductivity, phase fluctuations and the c -axis conductivity in bi-layer high temperature superconductors. *Phys. Rev. B* **65**, 024506 (2001).
14. Chaloupka, J., Bernhard, C. & Munzar, D. Microscopic gauge invariant theory of the c -axis infrared response of bi-layer cuprate superconductors and the origin of the superconductivity-induced absorption bands. *Phys. Rev. B* **79**, 184513 (2009).
15. Yu, L. *et al.* Evidence for two separate energy gaps in underdoped high-temperature cuprate superconductors from broadband infrared ellipsometry. *Phys. Rev. Lett.* **100**, 177004 (2008).
16. Ghiringhelli, G. *et al.* Long-range incommensurate charge fluctuations in $(\text{Y,Nd})\text{Ba}_2\text{Cu}_3\text{O}_{6+x}$. *Science* **337**, 821–825 (2012).
17. Blanco-Canosa, S. *et al.* Resonant X-ray scattering study of charge density wave correlations in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. *Phys. Rev. B* **90**, 054513 (2014).
18. Chang, J. *et al.* Direct observation of competition between superconductivity and charge density wave order in $\text{YBa}_2\text{Cu}_3\text{O}_{6.67}$. *Nature Phys.* **8**, 871–876 (2012).

Acknowledgements The research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ERC Grant Agreement no. 319286 (QMAC). Funding from the priority program SFB925 of the German Science Foundation is acknowledged. Portions of this research were carried out at the Linac Coherent Light Source (LCLS) at the SLAC National Accelerator Laboratory. The LCLS is an Office of Science User Facility operated for the US Department of Energy Office of Science by Stanford University. This work was supported by the Swiss National Supercomputing Centre under project ID s404. This work was supported by the Swiss National Science Foundation through its National Centre of Competences in Research MUST.

Author Contributions A.C. conceived this project. R.M. and M. Först led the diffraction experiment, supported by S.O.M., M.C., H.T.L., J.S.R., J.M.G., M.P.M. and A.F. R.M. and A.S. analysed the data. A.S. performed the DFT calculations, with support from A.G., M. Fechner and N.A.S. The sample was grown by T.L. and B.K. R.M. and A.C. wrote the manuscript, with feedback from all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.C. (andrea.cavalleri@mpsd.mpg.de) or R.M. (roman.mankowsky@mpsd.mpg.de).

METHODS

Experimental details. The X-ray diffraction measurements were carried out with 6.7 keV pulses at the X-ray pump probe (XPP) beamline of the LCLS. The energy of the X-rays was selected using a channel-cut Si (111) monochromator with a resolution of 1 eV. The diffraction from each pulse was recorded individually without averaging using a diode. Shot-to-shot normalization to the intensity monitor after the monochromator was used to correct the detected signals for intensity and wavelength fluctuations of the X-ray pulses. The experiment was carried out in grazing-incidence geometry with an angle of 5° between the X-rays and the sample surface.

The $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$ sample was excited with mid-infrared pulses of ~ 300 fs duration, generated by optical parametric down-conversion and difference-frequency generation of near-infrared pulses from a titanium-sapphire laser. These pulses were tuned to 15 μm wavelength with 2 μm bandwidth, chosen to be in resonance with the B_{1u} phonon mode. The measurement was carried out at a repetition rate of 120 Hz, while the repetition rate of the mid-infrared pulses was set to 60 Hz. This allowed us to measure the equilibrium and excited states for each time delay, to correct for any drifts of the free-electron laser.

Structure factor calculations. To deduce the amplitudes Q_i of the atomic displacements along the eigenvectors \mathbf{e}_i of the A_g coordinates from the changes in scattered intensity $I \propto |F|^2$, we calculated the corresponding modulation of the structure factors F . Specifically, we quantified $F = \sum_j f_j \exp(-i\mathbf{G} \cdot \mathbf{r}_j)$, where \mathbf{G} is the reciprocal

lattice vector of the corresponding diffraction peak, f_j are the atomic scattering factors and \mathbf{r}_j is the position of the j th atom in the unit cell. By calculating the structure factors for the equilibrium atomic positions $\mathbf{r}_j = \mathbf{r}_j^{(0)}$ and the transient structure $\mathbf{r}_j' = \mathbf{r}_j^{(0)} + \sum_i Q_i \mathbf{e}_{ij}$, the relative change in diffracted intensity was evaluated as

$$\Delta I = \left(|F(\mathbf{r}_j')|^2 - |F(\mathbf{r}_j)|^2 \right) / |F(\mathbf{r}_j)|^2.$$

The changes in signal amplitude are calculated for A_g amplitudes Q_i as predicted by DFT calculations for a certain infrared amplitude $Q_{B_{1u}}$ and compared with the experimental findings to determine the quantitative values.

DFT calculations. The phonon modes and the nonlinear phonon couplings were obtained using DFT calculations with plane-wave basis sets and projector augmented-wave pseudopotentials^{19,20} as implemented in the VASP software package²¹. The local-density approximation was used for the exchange and correlations. We used a cut-off energy of 950 eV for plane-wave expansion and a $4 \times 8 \times 4$ k -point grid for the Brillouin zone integration in the self-consistent cycles. We used the experimental lattice parameters of the $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$ ortho-II structure, but relaxed the internal coordinates. The interatomic force constants were calculated using the frozen-phonon method²², and the PHONOPY software package was used to calculate the phonon frequencies and normal modes²³. After the normal modes were identified, the total energy was calculated as a function of the Q_{IR} and Q_{R} phonon mode amplitudes to obtain the energy surfaces. The nonlinear coupling between the infrared and Raman modes was obtained by fitting the energy surfaces to the polynomial

$$H = \frac{1}{2} \omega_{\text{IR}}^2 Q_{\text{IR}}^2 + \frac{1}{2} \omega_{\text{R}}^2 Q_{\text{R}}^2 - a_{21} Q_{\text{IR}}^2 Q_{\text{R}}$$

Coupling strengths of the A_g to the B_{1u} modes. The energy potential of an A_g Raman mode in the presence of a cubic nonlinear coupling to an infrared mode is $V_{\text{R}} = \omega_{\text{R}}^2 Q_{\text{R}}^2 / 2 - a_{21} Q_{\text{IR}}^2 Q_{\text{R}}$. At equilibrium, when the infrared mode is not excited, the potential of the Raman mode has a minimum at $Q_{\text{R}} = 0$ because the structure is stable at equilibrium. However, when the infrared mode is excited externally, the minimum of the Raman mode shifts by an amount $a_{21} Q_{\text{IR}}^2 / \omega_{\text{R}}^2$. The minima of the A_g modes energy potentials as obtained from DFT calculations for a frozen displacement of the B_{1u} mode of 0.14 $\text{\AA}/u$ are reported in Extended Data Table 1. The mode displacements are given in terms of the amplitude Q_i of the dimensionless eigenvectors of the mode and have units of $\text{\AA}/u$, where u is the atomic mass unit. The atomic displacement due to an amplitude Q_i of a mode is given by $U_j = (Q_i / \sqrt{m_j}) \mathbf{e}_{ij}$, where U_j is the displacement of the j th atom, m_j is the mass of this atom and \mathbf{e}_{ij} is the corresponding component of the normal-mode vector. We note that \mathbf{e}_{ij} is normalized and dimensionless.

Transient crystal structure. The static crystal structure of $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$ (ortho-II) is given in Extended Data Table 2. The lattice constants are $a = 7.6586$ \AA , $b = 3.8722$ \AA and $c = 11.725$ \AA , as determined by single-crystal X-ray diffraction at 100 K. The light-induced displacements of the atomic positions at peak change in diffracted intensity are reported in Extended Data Table 2.

Changes in the electronic structure. The changes in the electronic structure due to the light-induced distortions were studied using the generalized full-potential method within the local-density approximation as implemented in the WIEN2k package²⁴. Muffin-tin radii of 2.35, 2.5, 1.78 and 1.53 Bohr radii (one Bohr radius equals 0.529 \AA) were used for Y, Ba, Cu and O, respectively, and a $20 \times 40 \times 20$

k -point grid was used for the Brillouin zone integration. The plane-wave cut-off was set such that $RK_{\text{max}} = 7.0$, where K_{max} is the plane-wave cut-off and R is the small muffin-tin radius, that is, 1.53 Bohr radii. The density of states was generated with a $32 \times 64 \times 32$ k -point grid. Calculations are presented for the equilibrium structure and the transient displaced structure for three B_{1u} amplitudes: the amplitude 0.3 $\text{\AA}/u$ determined here, the amplitude 0.8 $\text{\AA}/u$ estimated in refs 9, 10, and a larger amplitude of 1.2 $\text{\AA}/u$.

Our calculated electronic structure of the equilibrium $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$ (ortho-II), shown in Fig. 2, is similar to the one calculated previously^{25,26}. The bands near the Fermi level are derived from the Cu 3d states from both the planes and the chains. The four bands that have high dispersion along the path X–S–Y– Γ are due to the planar Cu $d_{x^2-y^2}$ states. The broad band that has very little dispersion along S–Y is due to the Cu d_{z^2} states from the filled chain. The O-deficient chains that control the hole doping give rise to fairly flat electronic bands with dominant Cu d_{xz} and d_{yz} character that are very close to the Fermi level at the Y point. The electronic structure calculations predict some hybridization between these bands and the planar Cu bands, which creates an anticrossing near Y. In the equilibrium structure, this anticrossing is close to the Fermi level, giving rise to pockets with unfilled-chain Cu character in the Fermi surface (Extended Data Figs 4 and 5).

The displacements due to the nonlinear couplings cause noticeable changes to the electronic structure around the Fermi level. There are three main effects.

(i) The light-induced displacements reduce the width of the planar Cu bands, which leads to an increase in the planar Cu contribution to the density of states at the Fermi level.

(ii) The atomic displacements cause a transfer of charge from the planes to the O-deficient chains. As the unfilled Cu chain bands decrease in energy and move below the Fermi level with increasing light-induced displacements, the planar Cu states increase in energy, becoming less occupied. That is, there is an effective hole doping of the planar Cu states owing to the light-induced displacements (Extended Data Fig. 6).

(iii) The changes in the relative occupations of the bands also cause a topological change in the Fermi surface. The light-induced displacements increase the filling of the unfilled-chain Cu bands, which decreases the size of the pockets in the Fermi surface. Above a threshold $Q_{B_{1u}}$ amplitude of 0.8 $\text{\AA}/u$, the O-deficient chain Cu bands become fully filled and the Fermi surface consists solely of two-dimensional planar Cu sheets and one-dimensional filled-chain Cu sheets.

Quartic-order coupling. To verify that the nonlinear phonon coupling is dominated by the third-order contribution, as discussed in the main text, we checked for signals at the next (fourth) order, described by the term $Q_{\text{IR}}^2 Q_j^2$ in the nonlinear Hamiltonian

$$H_{\text{NL}} = \frac{1}{2} \omega_{\text{R}}^2 Q_j^2 - a_{21} Q_{\text{IR}}^2 Q_j - a_{22} Q_{\text{IR}}^2 Q_j^2$$

As noted in the text, when the directly driven infrared mode is of B_{1u} symmetry, the only modes to which there is non-zero third-order coupling are those of A_g symmetry. However, coupling to any mode Q_j , in particular to in-plane B_g modes, is allowed through $Q^2 Q^2$ coupling.

We note first that small-amplitude B_{1u} excitations would simply renormalize the frequency of a second mode Q_j . This can be directly deduced from the equation of motion, where the driving force is given by the coupling term $a_{22} Q_{\text{IR}}^2 Q_j$, which is linear in Q_j :

$$\ddot{Q}_j + 2\gamma_j \dot{Q}_j + \omega_j^2 Q_j = 2a_{22} Q_{\text{IR}}^2 Q_j$$

On Q_{IR} displacement, the anharmonically coupled mode experiences a renormalization of its frequency: $\omega_j' = \omega_j \sqrt{1 - 2a_{22} Q_{\text{IR}}^2}$.

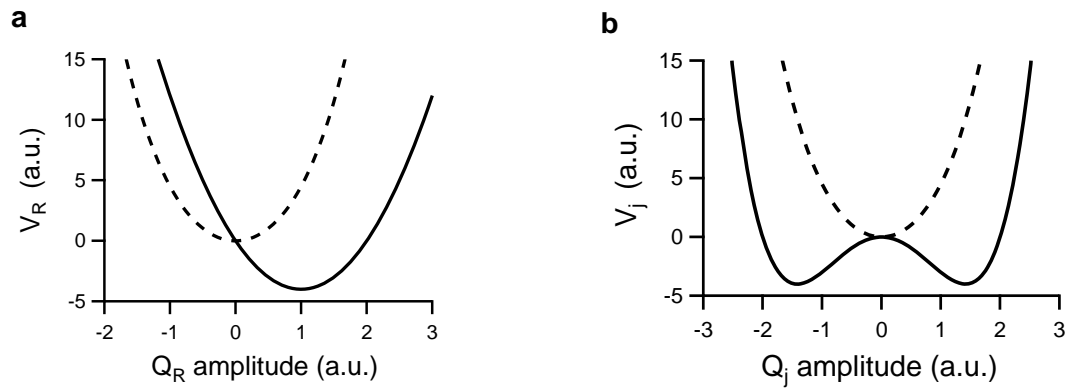
However, above a threshold amplitude Q_{IR}^* , the frequency of the second mode Q_j becomes imaginary³ and the lattice becomes unstable. Importantly, such instability can take place in two directions, depending on the random instantaneous state of the system (mode amplitude Q_j and its velocity dQ_j/dt). This manifests in a change from a parabolic to a double-well energy potential as shown in Extended Data Fig. 1.

Hence, fourth-order effects need to be identified by analysing the diffraction of each individual X-ray pulse, whereas the unsorted average is expected to be zero even if the quartic coupling is sizeable. In the experiment, we sorted all positive and negative deviations from the average signal of all shots to obtain the $Q^2 Q^2$ response at a specific time delay. Averaging them separately and subtracting negative deviations from positive then gives the intensity changes from $Q^2 Q^2$ only.

Time-resolved X-ray diffraction was measured for four Bragg reflections, sensitive to A_g and to B_{2g} displacements. The results of these experiments are shown in Extended Data Fig. 2.

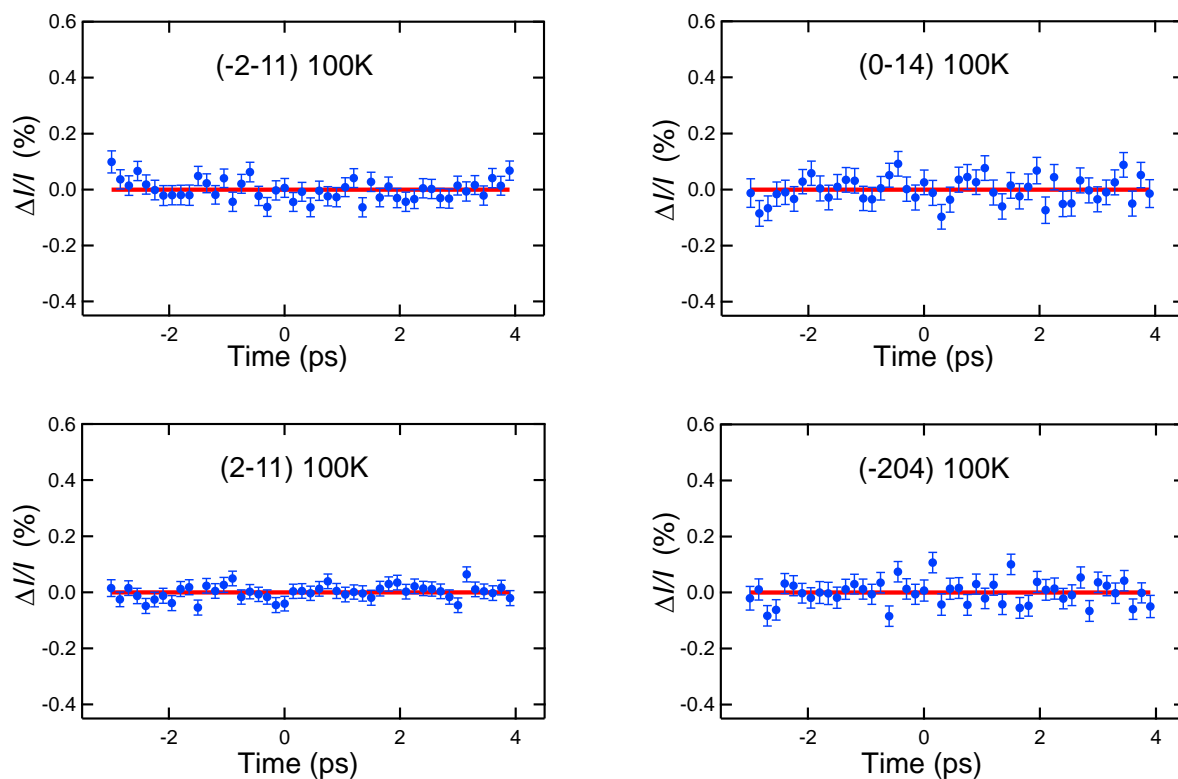
Within our resolution, we find no evidence of quartic contributions. The amplitude of the infrared motion is below the threshold beyond which fourth-order coupling induces lattice displacements.

19. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953 (1994).
20. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758 (1999).
21. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169 (1996).
22. Parlinski, K., Li, Z. & Kawazoe, Y. First-principles determination of the soft mode in cubic ZrO_2 . *Phys. Rev. Lett.* **78**, 4063 (1997).
23. Togo, A., Oba, F. & Tanaka, I. First-principles calculations of the ferroelastic transition between rutile-type and CaCl_2 -type SiO_2 at high pressures. *Phys. Rev. B* **78**, 134106 (2008).
24. Blaha, P., Schwarz, K., Madsen, G., Kvasnicka, D. & Luitz, J. WIEN2k, <http://www.wien2k.at/> (2001).
25. Carrington, A. & Yelland, E. A. Band-structure calculations of Fermi-surface pockets in ortho-II $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$. *Phys. Rev. B* **76**, 140508(R) (2007).
26. Elfimov, I. S., Sawatzky, G. A. & Damascelli, A. Theory of Fermi-surface pockets and correlation effects in underdoped $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$. *Phys. Rev. B* **77**, 060504(R) (2008).



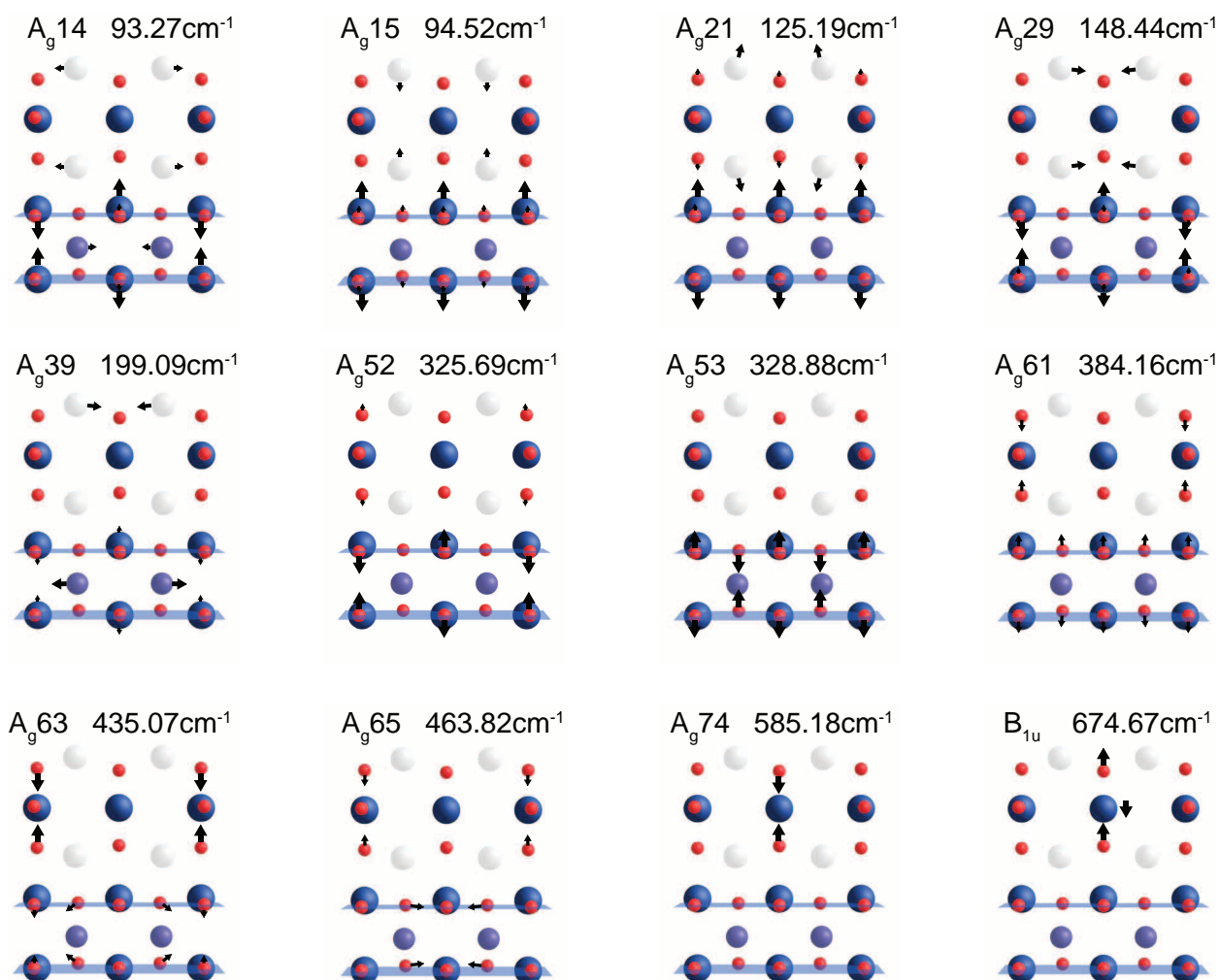
Extended Data Figure 1 | Nonlinear lattice dynamics in the limit of cubic and quartic coupling. Dashed lines: potential energy of a mode Q_R as a function of mode amplitude. **a**, A static distortion Q_{IR}^* shifts the potential of all modes Q_R that are coupled through a $Q_{IR}^2 Q_R$ coupling (solid line), displacing

the equilibrium position towards a new minimum. **b**, Owing to quartic $Q_{IR}^2 Q_j^2$ coupling, the energy potential of a coupled mode Q_j is deformed symmetrically on static distortion Q_{IR}^* . The frequency of the mode first softens until it is destabilized, which manifests in a double well potential (solid line).

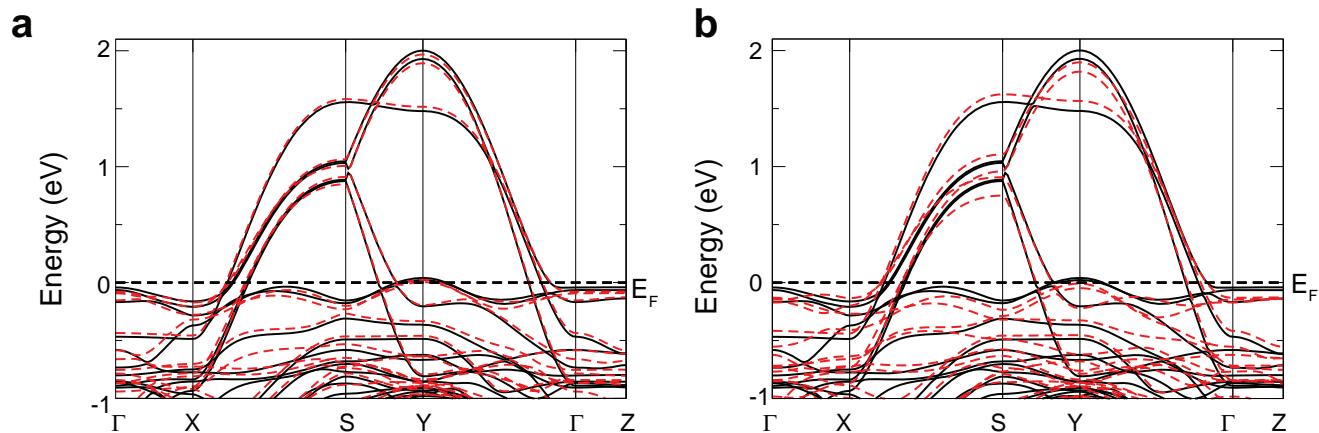


Extended Data Figure 2 | Changes in diffracted intensity of specific Bragg reflections from fourth-order coupling for different time delays between pump and probe pulse. We find no evidence of lattice distortions originating

from fourth-order contributions to the phonon coupling. The amplitude of the infrared mode Q_{IR} is below the threshold beyond which fourth-order effects destabilize coupled phonon modes. Error bars, 1σ (67% confidence interval).

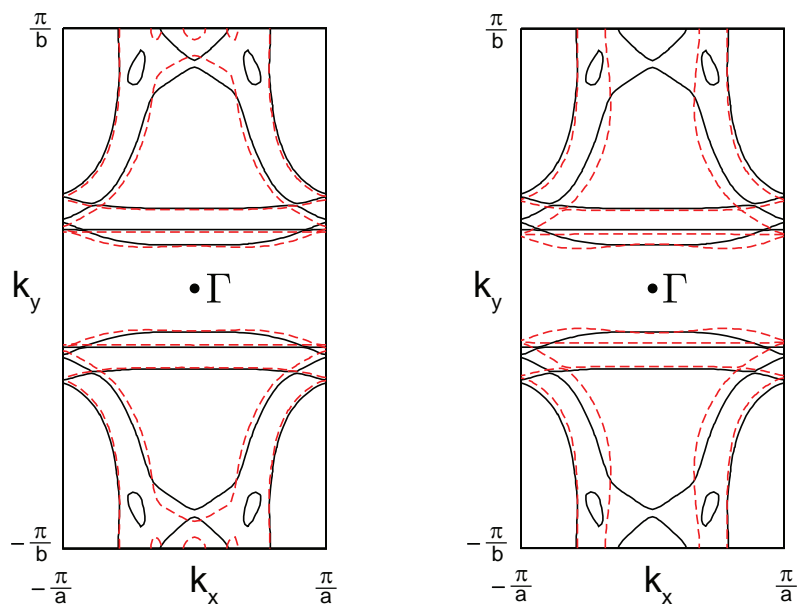


Extended Data Figure 3 | Phonon modes of ortho-II $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$. Sketches of the resonantly excited B_{1u} mode and all 11 A_g modes for which the coupling strengths (Extended Data Table 1) have been calculated.



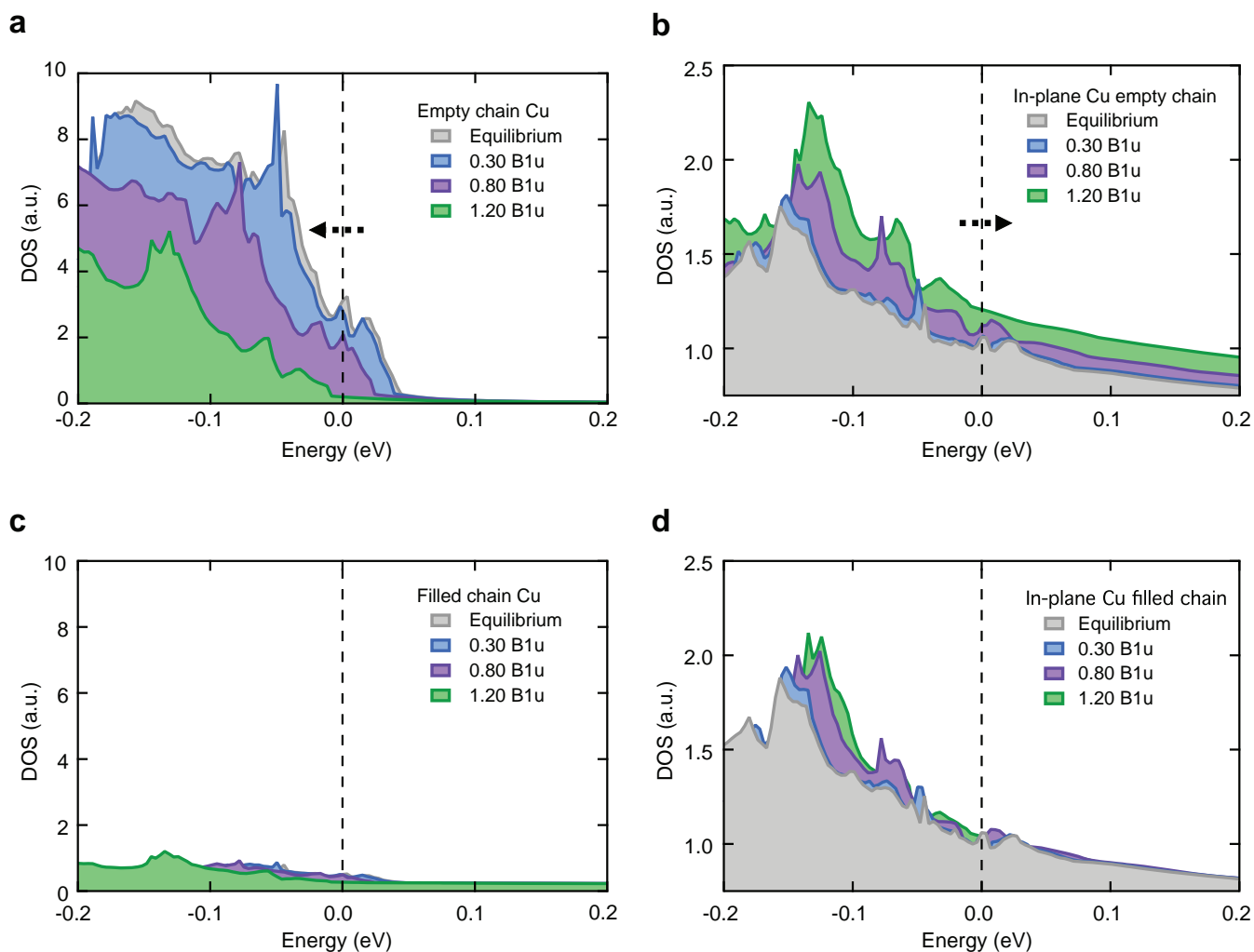
Extended Data Figure 4 | Band structure of the equilibrium (black line) and transient crystal structure. The band structure is plotted along $\Gamma(0, 0, 0) \rightarrow X(0.5, 0, 0) \rightarrow S(0.5, 0.5, 0) \rightarrow Y(0, 0.5, 0) \rightarrow \Gamma(0, 0, 0) \rightarrow$

$Z(0, 0, 0.5)$ for transient displaced structures corresponding to $Q_{B_{1u}}$ amplitudes of $0.8 \text{ \AA}/u$ (a), which is the amplitude estimated for the geometry of refs 9, 10, and $1.2 \text{ \AA}/u$ (b).



Extended Data Figure 5 | Cuts of the Fermi surface of the equilibrium (black line) and transient crystal structures (red dashed line) at $k_z = 0$. In the equilibrium structure, the bands of the unfilled-chain Cu atoms give rise to pockets in the Fermi surface. The light-induced displacements shift the densities of states of these bands to lower energies, increasing the filling and

reducing the pockets. Above a threshold of $0.8 \text{ \AA}/\sqrt{u}$, the O-deficient chain bands become fully filled, the pockets close and the Fermi surface consists solely of two-dimensional planar Cu sheets and one-dimensional filled-chain states. The Fermi surface is shown in the displaced state for Q_{B1u} amplitudes of $0.8 \text{ \AA}/\sqrt{u}$ (left) and $1.2 \text{ \AA}/\sqrt{u}$ (right).



Extended Data Figure 6 | Changes in the density of states in the CuO₂ plane and the Cu–O chains. These are obtained from a projection of the density of states onto the copper muffin-tin spheres. **a, b**, In the light-induced state, the density of states of the O-deficient chain lowers in energy (**a**), whereas the opposite effect is observed for the Cu in the plane below (**b**). This corresponds to charge transfer from the planes to the chains. **c, d**, The density of states of the

filled chain Cu is not strongly affected (**c**). The bands of the planar Cu atoms narrow, which leads to an increase in the density of states near the Fermi level both at sites with filled (**d**) and empty chains (**b**). The effect is already visible for a Q_{B1u} amplitude of 0.3 \AA/u (blue) but becomes more prominent for larger displacements of 0.8 \AA/u (purple) and 1.2 \AA/u (green).

Extended Data Table 1 | Mode displacements

Mode	Displacement ($\text{\AA}\sqrt{u}$)
Ag14	-0.002
Ag15	0.031
Ag21	-0.038
Ag29	-0.023
Ag39	0.000
Ag52	0.007
Ag53	0.000
Ag61	-0.007
Ag63	0.007
Ag65	-0.001
Ag74	0.020

Energy potential minima of the A_g modes as obtained from DFT calculations for a frozen displacement of the B_{1u} mode of $0.14 \text{ \AA}/u$, which corresponds to a change in apical O–Cu distance of 2.2 pm.

Extended Data Table 2 | Equilibrium structure of $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$ and light-induced displacements

Atom	Equilibrium Structure (\AA)			Displacements (pm) $0.3\text{\AA}/u$			Displacements (pm) $0.8\text{\AA}/u$			Displacements (pm) $1.2\text{\AA}/u$		
	x	y	z	x	y	z	x	y	z	x	y	z
Y	1.922	1.936	5.863	0.111	0.000	0.000	0.769	0.000	0.000	1.835	0.000	0.000
Y	5.737	1.936	5.863	-0.111	0.000	0.000	-0.769	0.000	0.000	-1.835	0.000	0.000
Ba	1.861	1.936	2.224	0.159	0.000	0.027	0.142	0.000	0.040	0.167	0.000	0.227
Ba	5.797	1.936	9.501	-0.159	0.000	-0.027	-0.142	0.000	-0.040	-0.167	0.000	-0.227
Ba	1.861	1.936	9.501	0.159	0.000	-0.027	0.142	0.000	-0.040	0.167	0.000	-0.227
Ba	5.797	1.936	2.224	-0.159	0.000	0.027	-0.142	0.000	0.040	-0.167	0.000	0.227
Cu	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cu	3.829	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cu	0.000	0.000	4.227	0.000	0.000	-0.376	0.000	0.000	-0.799	0.000	0.000	-1.659
Cu	0.000	0.000	7.498	0.000	0.000	0.376	0.000	0.000	0.799	0.000	0.000	1.659
Cu	3.829	0.000	4.231	0.000	0.000	-1.032	0.000	0.000	-4.999	0.000	0.000	-11.451
Cu	3.829	0.000	7.494	0.000	0.000	1.032	0.000	0.000	4.999	0.000	0.000	11.451
O	0.000	1.936	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
O	1.915	0.000	4.443	-0.055	0.000	0.066	-0.233	0.000	0.293	-1.024	0.000	1.144
O	5.743	0.000	7.282	0.055	0.000	-0.066	0.233	0.000	-0.293	1.024	0.000	-1.144
O	1.915	0.000	7.282	-0.055	0.000	-0.066	-0.233	0.000	-0.293	-1.024	0.000	-1.144
O	5.743	0.000	4.443	0.055	0.000	0.066	0.233	0.000	0.293	1.024	0.000	1.144
O	0.000	1.936	4.431	0.000	0.000	-0.015	0.000	0.000	-0.076	0.000	0.000	0.215
O	0.000	1.936	7.294	0.000	0.000	0.015	0.000	0.000	0.076	0.000	0.000	-0.215
O	3.829	1.936	4.440	0.000	0.000	0.127	0.000	0.000	0.490	0.000	0.000	1.690
O	3.829	1.936	7.285	0.000	0.000	-0.127	0.000	0.000	-0.490	0.000	0.000	-1.690
O	0.000	0.000	1.857	0.000	0.000	-0.057	0.000	0.000	0.199	0.000	0.000	-0.382
O	0.000	0.000	9.868	0.000	0.000	0.057	0.000	0.000	-0.199	0.000	0.000	0.382
O	3.829	0.000	1.758	0.000	0.000	1.335	0.000	0.000	7.355	0.000	0.000	12.342
O	3.829	0.000	9.967	0.000	0.000	-1.335	0.000	0.000	-7.355	0.000	0.000	-12.342

The light-induced atomic displacements are given for a $Q_{B_{1u}}$ amplitude of 0.3, 0.8 and $1.2\text{\AA}/u$.

Single-shot compressed ultrafast photography at one hundred billion frames per second

Liang Gao^{1*}, Jinyang Liang^{1*}, Chiye Li¹ & Lihong V. Wang¹

The capture of transient scenes at high imaging speed has been long sought by photographers^{1–4}, with early examples being the well known recording in 1878 of a horse in motion⁵ and the 1887 photograph of a supersonic bullet⁶. However, not until the late twentieth century were breakthroughs achieved in demonstrating ultrahigh-speed imaging (more than 10^5 frames per second)⁷. In particular, the introduction of electronic imaging sensors based on the charge-coupled device (CCD) or complementary metal–oxide–semiconductor (CMOS) technology revolutionized high-speed photography, enabling acquisition rates of up to 10^7 frames per second⁸. Despite these sensors' widespread impact, further increasing frame rates using CCD or CMOS technology is fundamentally limited by their on-chip storage and electronic readout speed⁹. Here we demonstrate a two-dimensional dynamic imaging technique, compressed ultrafast photography (CUP), which can capture non-repetitive time-evolving events at up to 10^{11} frames per second. Compared with existing ultrafast imaging techniques, CUP has the prominent advantage of measuring an x – y – t (x , y , spatial coordinates; t , time) scene with a single camera snapshot, thereby allowing observation of transient events with temporal resolution as tens of picoseconds. Furthermore, akin to traditional photography, CUP is receive-only, and so does not need the specialized active illumination required by other single-shot ultrafast imagers^{2,3}. As a result, CUP can image a variety of luminescent—such as fluorescent or bioluminescent—objects. Using CUP, we visualize four fundamental physical phenomena with single laser shots only: laser pulse reflection and refraction, photon racing in two media, and faster-than-light propagation of non-information (that is, motion that appears faster than the speed of light but cannot convey information). Given CUP's capability, we expect it to find widespread applications in both fundamental and applied sciences, including biomedical research.

To record events occurring at subnanosecond scale, currently the most effective approach is to use a streak camera, that is, an ultrafast photo-detection system that transforms the temporal profile of a light signal into a spatial profile by shearing photoelectrons perpendicularly to their direction of travel with a time-varying voltage¹⁰. However, a typical streak camera is a one-dimensional imaging device—a narrow entrance slit (10 – $50\text{ }\mu\text{m}$ wide) limits the imaging field of view (FOV) to a line. To achieve two-dimensional (2D) imaging, the system thus requires additional mechanical or optical scanning along the orthogonal spatial axis. Although this paradigm is capable of providing a frame rate fast enough to catch photons in motion^{11,12}, the event itself must be repetitive—following exactly the same spatiotemporal pattern—while the entrance slit of a streak camera steps across the entire FOV. In cases where the physical phenomena are either difficult or impossible to repeat, such as optical rogue waves¹³, nuclear explosion, and gravitational collapse in a supernova, this 2D streak imaging method is inapplicable.

To overcome this limitation, here we present CUP (Fig. 1), which can provide 2D dynamic imaging using a streak camera without employing any mechanical or optical scanning mechanism with a single exposure. On the basis of compressed sensing¹⁴, CUP works by encoding the spatial domain with a pseudo-random binary pattern, followed by a shearing

operation in the temporal domain, performed using a streak camera with a fully opened entrance slit. This encoded, sheared three-dimensional (3D) x , y , t scene is then measured by a 2D detector array, such as a CCD, with a single snapshot. The image reconstruction process follows a strategy similar to compressed-sensing-based image restoration^{15–19}—iteratively estimating a solution that minimizes an objective function.

By adding a digital micromirror device as the spatial encoding module and applying the CUP reconstruction algorithm, we transformed a conventional one-dimensional streak camera into a 2D ultrafast imaging device. The resultant system can capture a single, non-repetitive event at up to 100 billion frames per second with appreciable sequence depths (up to 350 frames per acquisition). Moreover, by using a dichroic mirror to separate signals into two colour channels, we expand CUP's functionality into the realm of four dimensional (4D) x , y , t , λ ultrafast imaging, maximizing the information content that we can simultaneously acquire from a single instrument (Methods).

CUP operates in two steps: image acquisition and image reconstruction. The image acquisition can be described by a forward model (Methods). The input image is encoded with a pseudo-random binary pattern and then temporally dispersed along a spatial axis using a streak camera. Mathematically, this process is equivalent to successively applying a spatial encoding operator, C , and a temporal shearing operator, S , to the intensity distribution from the input dynamic scene, $I(x, y, t)$:

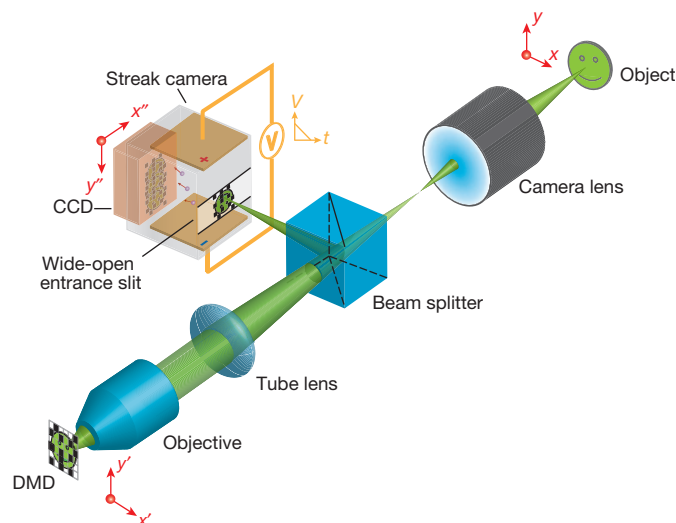


Figure 1 | CUP system configuration. CCD, charge-coupled device; DMD, digital micromirror device; V , sweeping voltage; t , time. Since each micromirror ($7.2\text{ }\mu\text{m} \times 7.2\text{ }\mu\text{m}$) of the DMD is much larger than the light wavelength, the diffraction angle is small ($\sim 4^\circ$). With a collecting objective of numerical aperture $\text{NA} = 0.16$, the throughput loss caused by the DMD's diffraction is negligible. See text for details of operation. Equipment details: camera lens, Fujinon CF75HA-1; DMD, Texas Instruments DLP LightCrafter; microscope objective, Olympus UPLSAPO 4 \times ; tube lens, Thorlabs AC254-150-A; streak camera, Hamamatsu C7700; CCD, Hamamatsu ORCA-R2.

¹Optical Imaging Laboratory, Department of Biomedical Engineering, Washington University in St Louis, Campus Box 1097, One Brookings Drive, St Louis, Missouri 63130, USA.

*These authors contributed equally to this work.

$$I_s(x'', y'', t) = SCI(x, y, t) \quad (1)$$

where $I_s(x'', y'', t)$ represents the resultant encoded, sheared scene. Next, I_s is imaged by a CCD, a process that can be mathematically formulated as:

$$E(m, n) = TI_s(x'', y'', t) \quad (2)$$

Here, T is a spatiotemporal integration operator (spatially integrating over each CCD pixel and temporally integrating over the exposure time). $E(m, n)$ is the optical energy measured at pixel m, n on the CCD. Substituting equation (1) into equation (2) yields

$$E(m, n) = OI(x, y, t) \quad (3)$$

where O represents a combined linear operator, that is, $O = TSC$.

The image reconstruction is solving the inverse problem of equation (3). Given the operator O and spatiotemporal sparsity of the event, we can reasonably estimate the input scene, $I(x, y, t)$, from measurement, $E(m, n)$, by adopting a compressed-sensing algorithm, such as two-step iterative shrinkage/thresholding (TwIST)¹⁶ (detailed in Methods). The reconstructed frame rate, r , is determined by:

$$r = \frac{v}{\Delta y''} \quad (4)$$

Here v is the temporal shearing velocity of the operator S , that is, the shearing velocity of the streak camera, and $\Delta y''$ is the CCD's binned pixel size along the temporal shearing direction of the operator S .

The configuration of the CUP is shown in Fig. 1. The object is first imaged by a camera lens with a focal length of 75 mm. The intermediate image is then passed to a DMD by a 4f imaging system consisting of a tube lens (focal length 150 mm) and a microscope objective (focal length 45 mm, numerical aperture 0.16). To encode the input image, a pseudo-random binary pattern is generated and displayed on the DMD, with a binned pixel size of $21.6 \mu\text{m} \times 21.6 \mu\text{m}$ (3×3 binning).

The light reflected from the DMD is collected by the same microscope objective and tube lens, reflected by a beam splitter, and imaged onto the entrance slit of a streak camera. To allow 2D imaging, this entrance slit is opened to its maximal width (~ 5 mm). Inside the streak camera, a sweeping voltage is applied along the y'' axis, deflecting the encoded image frames towards different y'' locations according to their times of arrival. The final temporally dispersed image is captured by a CCD (672×512 binned pixels (2×2 binning)) with a single exposure.

To characterize the system's spatial frequency responses, we imaged a dynamic scene, namely, a laser pulse impinging upon a stripe pattern with varying periods as shown in Fig. 2a. The stripe frequency (in line pairs per mm) descends stepwise along the x axis from one edge to the other. We shone a collimated laser pulse (532 nm wavelength, 7 ps pulse duration) on the stripe pattern at an oblique angle of incidence α of $\sim 30^\circ$ with respect to the normal of the pattern surface. The imaging system faced the pattern surface and collected the scattered photons from the scene. The impingement of the light wavefront upon the pattern surface was imaged by CUP at 100 billion frames per second with the streak camera's shearing velocity set to 1.32 mm ns^{-1} . The reconstructed 3D x, y, t image of the scene in intensity (W m^{-2}) is shown in Fig. 2b, and

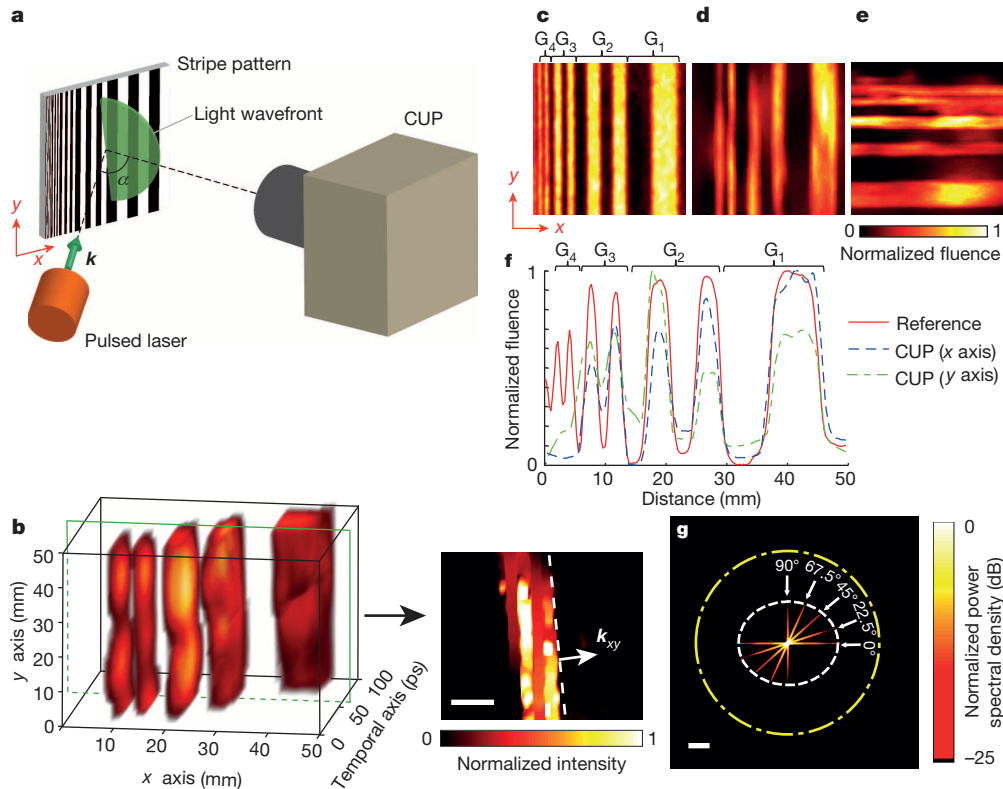


Figure 2 | Characterization of CUP's spatial frequency responses.

a, Experimental setup. A collimated laser pulse illuminates a stripe pattern obliquely. The CUP system faces the scene and collects the scattered photons. Laser is an Attodyne APL-4000. **b**, Left panel, reconstructed x, y, t datacube representing a laser pulse impinging upon the stripe pattern. Right panel, the representative frame for $t = 60$ ps, where the dashed line indicates the light wavefront, and the white arrow denotes the in-plane light propagation direction (k_{xy}). The entire temporal evolution of this event is shown in Supplementary Video 1. **c**, Reference image captured without introducing temporal dispersion (see **f** for explanation of G_4 – G_1). **d**, **e**, Projected vertical and

horizontal stripe images calculated by summing over the x, y, t datacube voxels along the temporal axis. **f**, Comparison of average light fluence distributions along the x axis from **c** (labelled as Reference) and **d** (labelled as CUP (x axis)), as well as along the y axis from **e** (labelled as CUP (y axis)). G_1, G_2, G_3 and G_4 refer to the stripe groups with ascending spatial frequencies. **g**, Spatial frequency responses for five different orientations of the stripe pattern. The inner white dashed ellipse represents the CUP system's band limit, and the outer yellow dashed-dotted circle represents the band limit of the optical modulation transfer function without temporal shearing. Scale bars: **b**, 10 mm; **g**, 0.1 mm^{-1} .

the corresponding time-lapse 2D x, y images ($50 \text{ mm} \times 50 \text{ mm}$ FOV; 150×150 pixels as nominal resolution) are provided in Supplementary Video 1.

Figure 2b also shows a representative temporal frame at $t = 60 \text{ ps}$. Within a 10 ps exposure, the wavefront propagates 3 mm in space. Including the thickness of the wavefront itself, which is $\sim 2 \text{ mm}$, the wavefront image is approximately 5 mm thick along the wavefront propagation direction. The corresponding intersection with the x - y plane is $5 \text{ mm}/\sin \alpha \approx 10 \text{ mm}$ thick, which agrees with the actual measurement ($\sim 10 \text{ mm}$).

We repeated the light sweeping experiment at four additional orientations of the stripe pattern (22.5° , 45° , 67.5° and 90° with respect to the x axis) and also directly imaged the scene without temporal dispersion to acquire a reference (Fig. 2c). We projected the x, y, t datacubes onto the x - y plane by summing over the voxels along the temporal axis. The resultant images at two representative angles (0° and 90°) are shown in Fig. 2d and e, respectively. We compare in Fig. 2f the average light fluence (J m^{-2}) distributions along the x axis from Fig. 2c and d as well as that along the y axis from Fig. 2e. The comparison indicates that CUP can recover spatial frequencies up to 0.3 line pairs per mm (groups G_1 , G_2 and G_3) along both x and y axes; however, the stripes in group G_4 —having a fundamental spatial frequency of 0.6 line pairs per mm—are beyond the resolution of the CUP system.

We further analysed the resolution by computing the spatial frequency spectra of the average light fluence distributions for all five orientations of the stripe pattern (Fig. 2g). Each angular branch appears continuous rather than discrete because the object has multiple stripe groups of varied frequencies and each has a limited number of periods. As a result, the spectra from the individual groups are broadened and overlapped. The CUP's spatial frequency response band is delimited by the inner white dashed ellipse, whereas the band purely limited by the optical modulation transfer function of the system without temporal shearing—derived from the reference image (Fig. 2c)—is enclosed by the outer yellow dash-dotted circle. The CUP resolutions along the x and y axes are ~ 0.43 and 0.36 line pairs per mm, respectively, whereas the un-sheared-system resolution is ~ 0.78 line pairs per mm. Here, resolution is defined as the noise-limited bandwidth at the 3σ threshold above the average background, where σ is the noise defined by the standard deviation of the background. The resolution anisotropy is attributed to the spatiotemporal mixing along the y axis only. Thus, CUP trades spatial resolution and resolution isotropy for temporal resolution.

To demonstrate CUP's 2D ultrafast imaging capability, we imaged three fundamental physical phenomena with single laser shots: laser pulse reflection and refraction, and racing of two pulses in different media (air and resin). It is important to mention that, unlike a previous study¹¹, here we truly recorded one-time events: only a single laser pulse was fired during image acquisition. In these experiments, to scatter light from the media to the CUP system, we evaporated dry ice into the light path in the air and added zinc oxide powder into the resin, respectively.

Figure 3a and b shows representative time-lapse frames of a single laser pulse reflected from a mirror in the scattering air and refracted at an air–resin interface, respectively. The corresponding movies are provided in Supplementary Videos 2 and 3. With a shearing velocity of 0.66 mm ns^{-1} in the streak camera, the reconstructed frame rate is 50 billion frames per second. Such a measurement allows the visualization of a single laser pulse's compliance with the laws of light reflection and refraction, the underlying foundations of optical science. It is worth noting that the heterogeneities in the images are probably attributable to turbulence in the vapour and non-uniform scattering in the resin.

To validate CUP's ability to quantitatively measure the speed of light, we imaged photon racing in real time. We split the original laser pulse into two beams: one beam propagated in the air and the other in the resin. The representative time-lapse frames of this photon racing experiment are shown in Fig. 3c, and the corresponding movie is provided in Supplementary Video 4. As expected, owing to the different refractive indices (1.0 in air and 1.5 in resin), photons ran faster in the air than in the resin. By tracing the centroid from the time-lapse laser pulse images

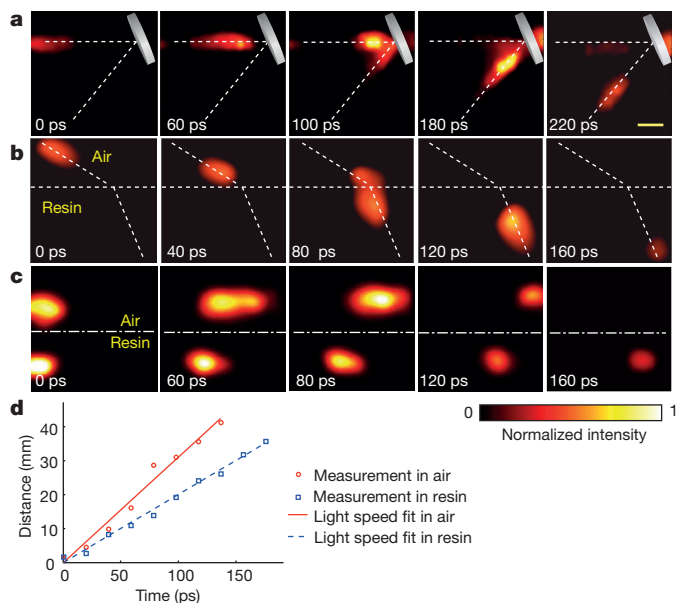


Figure 3 | CUP of laser pulse reflection, refraction, and racing in different media. **a**, Representative temporal frames showing a laser pulse being reflected from a mirror in air. The bright spot behind the mirror is attributed to imperfect resolution. **b**, As **a** but showing a laser pulse being refracted at an air–resin interface. **c**, As **a** but showing two racing laser pulses; one propagates in air and the other in resin. **d**, Recovered light speeds in air and in resin. The corresponding movies of these three events (**a**, **b**, **c**) are shown in Supplementary Videos 2, 3 and 4. Scale bar (in top right subpanel), 10 mm .

(Fig. 3d), the CUP-recovered light speeds in the air and in the resin were $(3.1 \pm 0.5) \times 10^8 \text{ m s}^{-1}$ and $(2.0 \pm 0.2) \times 10^8 \text{ m s}^{-1}$, respectively, consistent with the theoretical values ($3.0 \times 10^8 \text{ m s}^{-1}$ and $2.0 \times 10^8 \text{ m s}^{-1}$). Here the standard errors are mainly attributed to the resolution limits.

By monitoring the time-lapse signals along the laser propagation path in the air, we quantified CUP's temporal resolution. Because the 7 ps pulse duration is shorter than the frame exposure time (20 ps), the laser pulse was considered as an approximate impulse source in the time domain. We measured the temporal point-spread functions (PSFs) at different spatial locations along the light path imaged at 50 billion frames per second (20 ps exposure time), and their full widths at half maxima averaged 74 ps . Additionally, to study the dependence of CUP's temporal resolution on the frame rate, we repeated this experiment at 100 billion frames per second (10 ps exposure time) and re-measured the temporal PSFs. The mean temporal resolution was improved from 74 ps to 31 ps at the expense of signal-to-noise ratio. At a higher frame rate (that is, higher shearing velocity in the streak camera), the light signals are spread over more pixels on the CCD camera, reducing the signal level per pixel and thereby causing more potential reconstruction artefacts.

To explore CUP's potential application in modern physics, we imaged apparent faster-than-light phenomena in 2D movies. According to Einstein's theory of relativity, the propagation speed of matter cannot surpass the speed of light in vacuum because it would need infinite energy to do so. Nonetheless, if the motion itself does not transmit information, its speed can be faster than light²⁰. This phenomenon is referred to as faster-than-light propagation of non-information. To visualize this phenomenon with CUP, we designed an experiment using a setup similar to that shown in Fig. 2a. The pulsed laser illuminated the scene at an oblique angle of incidence of $\sim 30^\circ$, and CUP imaged the scene normally (0° angle). To facilitate the calculation of speed, we imaged a stripe pattern with a constant period of 12 mm (Fig. 4a).

The movement of a light wavefront intersecting this stripe pattern is captured at 100 billion frames per second with the streak camera's shearing velocity set to 1.32 mm ns^{-1} . Representative temporal frames and the corresponding movie are provided in Fig. 4b and Supplementary Video 5, respectively. As shown in Fig. 4b, the white stripes shown in

Fig. 4a are illuminated sequentially by the sweeping wavefront. The speed of this motion, calculated by dividing the stripe period by their lit-up time interval, is $v_{\text{FTL}} = 12 \text{ mm}/20 \text{ ps} = 6 \times 10^8 \text{ m s}^{-1}$, twice the speed of light in the air due to the oblique incidence of the laser beam (FTL, faster than light). As shown in Fig. 4c, although the intersected wavefront—the only feature visible to the CUP system—travels from location A to B faster than the light wavefront, the actual information is carried by the wavefront itself, and therefore its transmission velocity is still limited by the speed of light in the air.

Secure communication using CUP is possible because the operator **O** is built upon a pseudo-randomly generated code matrix sheared at a preset velocity. The encrypted scene can therefore be decoded only by recipients who are granted access to the decryption key. Using a DMD (instead of a pre-made mask) as the field encoding unit in CUP facilitates pseudo-random key generation and potentially allows the encoding pattern to be varied for each exposure transmission, thereby minimizing the impact of theft with a single key decryption on the overall information security. Furthermore, compared with other compressed-sensing-based secure communication methods for either a 1D signal or a 2D image^{21–23}, CUP operates on a 3D data set, allowing transient events to be captured and communicated at faster speed.

Although not demonstrated here, CUP can potentially be coupled to a variety of imaging modalities, such as microscopes and telescopes, allowing us to image transient events at scales from cellular organelles to galaxies. For example, in conventional fluorescence lifetime imaging microscopy (FLIM)²⁴, point scanning or line scanning is typically employed to achieve 2D fluorescence lifetime mapping. However, these scanning instruments cannot collect light from all elements of a data set in parallel. As a result, when measuring an image of $N_x \times N_y$ pixels, there is a loss of throughput by a factor of $N_x \times N_y$ (point scanning) or N_y

(line scanning). Additionally, scanning-based FLIM suffers from severe motion artefacts when imaging dynamic scenes, limiting its application to fixed or slowly varying samples. By integrating CUP with FLIM, we could accomplish parallel acquisition of a 2D fluorescence lifetime map within a single snapshot, thereby providing a simple solution to these long-standing problems in FLIM.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 June; accepted 17 October 2014.

- Fuller, P. W. W. An introduction to high speed photography and photonics. *Imaging Sci. J.* **57**, 293–302 (2009).
- Li, Z., Zgadzaj, R., Wang, X., Chang, Y.-Y. & Downer, M. C. Single-shot tomographic movies of evolving light-velocity objects. *Nature Commun.* **5**, 3085 (2014).
- Nakagawa, K. *et al.* Sequentially timed all-optical mapping photography (STAMP). *Nature Photon.* **8**, 695–700 (2014).
- Heshmat, B., Satat, G., Barsi, C. & Raskar, R. Single-shot ultrafast imaging using parallax-free alignment with a tilted lenslet array. *CLEO Sci. Innov.* http://dx.doi.org/10.1364/CLEO_SI.2014.STu3E.7 (2014).
- Munn, O. & Beach, A. A horse's motion scientifically determined. *Sci. Am.* **39**, 241 (1878).
- Mach, E. & Salcher, P. Photographische Fixierung der durch Projectile in der Luft eingeleiteten Vorgänge. *Ann. Phys.* **268**, 277–291 (1887).
- Nebeker, S. Rotating mirror cameras. *High Speed Photogr. Photon. News.* **3**, 31–37 (1983).
- Kondo, Y. *et al.* Development of “HyperVision HPV-X” high-speed video camera. *Shimadzu Rev.* **69**, 285–291 (2012).
- El-Desouki, M. *et al.* CMOS image sensors for high speed applications. *Sensors* **9**, 430–444 (2009).
- Guide to Streak Cameras* http://www.hamamatsu.com/resources/pdf/sys/e_streakh.pdf (Hamamatsu Corp., Hamamatsu City, Japan, 2002).
- Velten, A., Lawson, E., Bardagjy, A., Bawendi, M. & Raskar, R. Slow art with a trillion frames per second camera. *Proc. SIGGRAPH* **44** (2011).
- Velten, A. *et al.* Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Commun.* **3**, 745 (2012).
- Solli, D. R., Ropers, C., Koonath, P. & Jalali, B. Optical rogue waves. *Nature* **450**, 1054–1057 (2007).
- Eldar, Y. C. & Kutyniok, G. *Compressed Sensing: Theory and Applications* (Cambridge Univ. Press, 2012).
- Figueiredo, M. A. T., Nowak, R. D. & Wright, S. J. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE J. Sel. Top. Signal Process.* **1**, 586–597 (2007).
- Bioucas-Dias, J. M. & Figueiredo, M. A. T. A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Process.* **16**, 2992–3004 (2007).
- Afonso, M. V., Bioucas-Dias, J. M. & Figueiredo, M. A. T. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. Image Process.* **19**, 2345–2356 (2010).
- Wright, S. J., Nowak, R. D. & Figueiredo, M. A. T. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Process.* **57**, 2479–2493 (2009).
- Afonso, M. V., Bioucas-Dias, J. M. & Figueiredo, M. A. T. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Image Process.* **20**, 681–695 (2011).
- Einstein, A. *Relativity: The Special and General Theory* (Penguin, 1920).
- Agrawal, S. & Vishwanath, S. Secrecy using compressive sensing. *Proc. Inform. Theory Workshop, IEEE*, 563–567 (2011).
- Orsdemir, A., Altun, H. O., Sharma, G. & Bocko, M. F. On the security and robustness of encryption via compressed sensing. *Proc. Military Commun. Conf., IEEE*, 1–7 (2008).
- Davis, D. L. Secure compressed imaging. US patent 5,907 619 (1999).
- Borst, J. W. & Visser, A. J. W. G. Fluorescence lifetime imaging microscopy in life sciences. *Meas. Sci. Technol.* **21**, 102002 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank N. Hagen for discussions and J. Ballard for a close reading of the manuscript. We also acknowledge Texas Instruments for providing the DLP device. This work was supported in part by National Institutes of Health grants DP1 EB016986 (NIH Director's Pioneer Award) and R01 CA186567 (NIH Director's Transformative Research Award).

Author Contributions L.G. built the system, performed the experiments, analysed the data and prepared the manuscript. J.L. performed some of the experiments, analysed the data and prepared the manuscript. C.L. prepared the sample and performed some of the experiments. L.V.W. contributed to the conceptual system, experimental design and manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.V.W. (lhwang@wustl.edu).

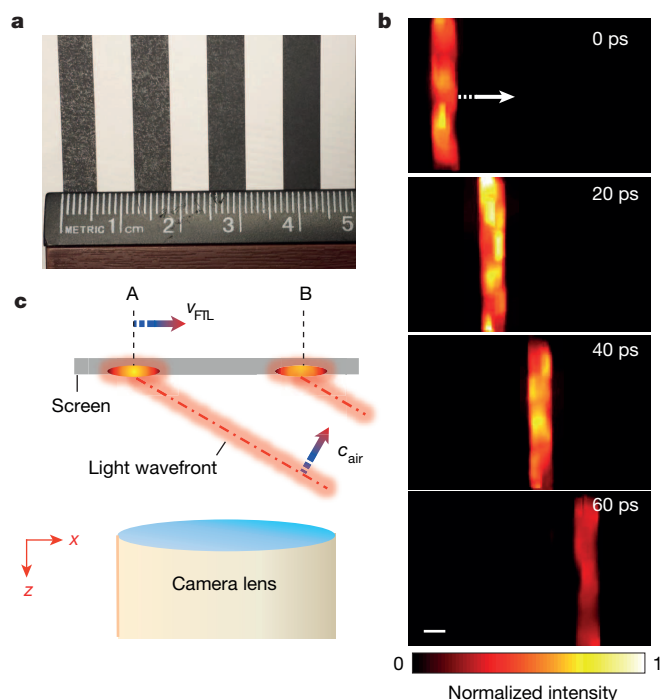


Figure 4 | CUP of faster-than-light propagation of non-information. **a**, Photograph of a stripe pattern with a constant period of 12 mm. **b**, Representative temporal CUP frames, showing the optical wavefront sweeping across the pattern. The corresponding movie is shown in Supplementary Video 5. The striped arrow indicates the motion direction. **c**, Illustration of the information transmission in this event. The intersected wavefront on the screen travels at a speed, v_{FTL} , which is twice the speed of light in air, c_{air} . However, because the light wavefront carries the actual information, the information transmission velocity from location A to B is still limited to c_{air} . Scale bar, 5 mm.

METHODS

Forward model. We describe CUP's image formation process using a forward model. The intensity distribution of the dynamic scene, $I(x, y, t)$, is first imaged onto an intermediate plane by an optical imaging system. Under the assumption of unit magnification and ideal optical imaging—that is, the PSF approaches a delta function—the intensity distribution of the resultant intermediate image is identical to that of the original scene. To encode this image, a mask which contains pseudo-randomly distributed, squared and binary-valued (that is, either opaque or transparent) elements is placed at this intermediate image plane. The image immediately after this encoding mask has the following intensity distribution:

$$I_c(x', y', t) = \sum_{i,j} I(x', y', t) C_{i,j} \text{rect}\left[\frac{x'}{d'} - (i + \frac{1}{2}), \frac{y'}{d'} - (j + \frac{1}{2})\right] \quad (5)$$

Here, $C_{i,j}$ is an element of the matrix representing the coded mask, i, j are matrix element indices, and d' is the mask pixel size. For each dimension, the rectangular function (rect) is defined as:

$$\text{rect}(x) = \begin{cases} 1, & \text{if } |x| \leq \frac{1}{2} \\ 0, & \text{else} \end{cases}$$

In this section, a mask or camera pixel is equivalent to a binned DMD or CCD pixel defined in the experiment.

This encoded image is then passed to the entrance port of a streak camera. By applying a voltage ramp, the streak camera acts as a shearing operator along the vertical axis (y'' axis in Extended Data Fig. 1) on the input image. If we again assume ideal optics with unit magnification, the sheared image can be expressed as

$$I_s(x'', y'', t) = I_c(x'', y'' - vt, t) \quad (6)$$

where v is the shearing velocity of the streak camera.

$I_s(x'', y'', t)$ is then spatially integrated over each camera pixel and temporally integrated over the exposure time. The optical energy, $E(m, n)$, measured at pixel m, n , is:

$$E(m, n) = \int dt \int dx'' \int dy'' I_s(x'', y'', t) \text{rect}\left[\frac{x''}{d''} - (m + \frac{1}{2}), \frac{y''}{d''} - (n + \frac{1}{2})\right] \quad (7)$$

Here, d'' is the camera pixel size. Accordingly, we can voxelize the input scene, $I(x, y, t)$, into $I_{i,j,k}$ as follows:

$$I(x, y, t) \approx \sum_{i,j,k} I_{i,j,k} \text{rect}\left[\frac{x}{d''} - (i + \frac{1}{2}), \frac{y}{d''} - (j + \frac{1}{2}), \frac{t}{\Delta_t} - (k + \frac{1}{2})\right] \quad (8)$$

where $\Delta_t = d''/v$. If the mask elements are mapped 1:1 to the camera pixels (that is, $d' = d''$) and perfectly registered, combining equations (5)–(8) yields:

$$E(m, n) = \frac{d'^{n+1}}{v} \sum_{k=0}^{n-1} C_{m,n-k} I_{m,n-k,k} \quad (9)$$

Here $C_{m,n-k} I_{m,n-k,k}$ represents a coded, sheared scene, and the inverse problem of equation (9) can be solved using existing compressed-sensing algorithms^{25–29}.

It is worth noting that only those indices where $n - k > 0$ should be included in equation (9). Thus, to convert equation (9) into a matrix equation, we need to augment the matrices C and I with an array of zeros. For example, to estimate a dynamic scene with dimensions $N_x \times N_y \times N_t$ (N_x, N_y and N_t are respectively the numbers of voxels along x, y and t), where the coded mask itself has dimensions $N_x \times N_y$, the actual matrices I and C used in equation (9) will have dimensions $N_x \times (N_y + N_t - 1) \times N_t$ and $N_x \times (N_y + N_t - 1)$, respectively, with zeros padded to the ends. After reconstruction, these extra voxels, containing non-zero values due to noise, are simply discarded.

CUP image reconstruction algorithm. Given prior knowledge of the coded mask matrix, to estimate the original scene from the CUP measurement, we need to solve the inverse problem of equation (9). This process can be formulated in a more general form as

$$\arg \min \left\{ \frac{1}{2} \|E - OI\|^2 + \beta \Phi(I) \right\} \quad (10)$$

where O is the linear operator, $\Phi(I)$ is the regularization function, and β is the regularization parameter. In CUP image reconstruction, we adopt an algorithm called two-step iterative shrinkage/thresholding (TwIST)²⁵, with $\Phi(I)$ in the form of total variation (TV):

$$\begin{aligned} \Phi(I) = & \sum_{k=0}^{N_t-1} \sum_{i=1}^{N_x \times N_y} \sqrt{(\Delta_i^h I_k)^2 + (\Delta_i^v I_k)^2} + \sum_{m=1}^{N_x} \sum_{i=1}^{N_y \times N_t} \sqrt{(\Delta_i^h I_m)^2 + (\Delta_i^v I_m)^2} \\ & + \sum_{n=1}^{N_y} \sum_{i=1}^{N_x \times N_t} \sqrt{(\Delta_i^h I_n)^2 + (\Delta_i^v I_n)^2}. \end{aligned} \quad (11)$$

Here we assume that the discretized form of I has dimensions $N_x \times N_y \times N_t$, and m, n, k are the three indices. I_m, I_n, I_k denote the 2D lattices along the dimensions m, n, k , respectively. Δ_i^h and Δ_i^v are horizontal and vertical first-order local difference operators on a 2D lattice. In TwIST, the minimization of the first term, $\|E - OI\|^2$, occurs when the actual measurement E closely matches the estimated measurement OI , while the minimization of the second term, $\beta \Phi(I)$, encourages I to be piecewise constant (that is, sparse in the gradient domain). The weighting of these two terms is empirically adjusted by the regularization parameter, β , to lead to the results that are most consistent with the physical reality. To reconstruct a datacube of size $150 \times 150 \times 350$ (x, y, t), approximately ten minutes is required on a computer with Intel i5-2500 CPU (3.3 GHz) and 8 GB RAM. The reconstruction process can be further accelerated by using GPUs.

Traditionally, the TwIST algorithm is initialized with a pseudo-random matrix as the discretized form of I and then converged to a solution by minimizing the objective function in equation (10). Thus no spatiotemporal information about the event is typically employed in the basic TwIST algorithm. However, it is important to remember that the solution of TwIST might not converge to a global minimum, and hence might not provide a physically reasonable estimate of the event. Therefore, the TwIST algorithm may include a supervision step that models the initial estimate of the event. For example, if the spatial or temporal range within which an event occurs is known a priori, one can assign non-zero values to only the corresponding voxels in the initial estimate of the discretized form of I and start optimization thereafter. Compared with the basic TwIST algorithm, the supervised-TwIST approach can substantially reduce reconstruction artefacts and therefore provide a more reliable solution.

Effective field-of-view measurement. In CUP, a streak camera (Hamamatsu C7700) temporally disperses the light. The streak camera's entrance slit can be fully opened to a 17 mm \times 5 mm rectangle (horizontal \times vertical axes). Without temporal dispersion, the image of this entrance slit on the CCD has an approximate size of 510×150 pixels. However, because of a small angle between each individual micro-mirror's on-state normal and the DMD's surface normal, the DMD as a whole needs to be tilted horizontally so that the incident light can be exactly retroreflected. With an NA of 0.16, the collecting objective's depth of focus thereby limits the horizontal encoding field of view (FOV) to approximately 150 pixels at the CCD. Extended Data Fig. 2 shows a temporally undispersed CCD image of the DMD mask. The effective encoded FOV is approximately 150×150 pixels. Note that with temporal dispersion, the image of this entrance slit on the CCD can be stretched along the y'' axis to approximately 150×500 pixels.

Calibration. To calibrate for operator matrix C , we use a uniform scene as the input image and apply zero sweeping voltage in the streak camera. The coded pattern on the DMD is therefore directly imaged onto the CCD without introducing temporal dispersion. A background image is also captured with all DMD pixels turned on. The illumination intensity non-uniformity is corrected for by dividing the coded pattern image by the background image pixel by pixel, yielding operator matrix C . Note that because CUP's image reconstruction is sensitive to mask misalignment, we use a DMD for better stability rather than pre-made masks that would require mechanical swapping between system alignment and calibration or data acquisition.

Limitation on CUP's frame rate and reconstructed datacube size. The CUP system's frame rate and temporal resolution are determined by the shearing velocity of the streak camera: a faster shearing velocity results in a higher frame rate and temporal resolution. Unless the illumination is intensified, however, the shortened observation time window causes the signal-to-noise ratio to deteriorate, which reduces image reconstruction quality. The shearing velocity thus should be balanced to accommodate a specific imaging application at a given illumination intensity.

In CUP, the size of the reconstructed event datacube, $N_x \times N_y \times N_t$, is limited fundamentally by the acceptance NA of the collecting objective, photon shot noise, and sensitivity of the photocathode tube, as well as practically by the number of binned CCD pixels ($N_R \times N_C$; N_R , the number of rows; N_C , the number of columns). Provided that the image formation closely follows the ideal forward model, the number of binned CCD pixels becomes the underlying limiting factor. Along the horizontal direction, the number of reconstructed voxels must be less than the number of detector columns, that is, $N_x \leq N_C$. In multicolour CUP, this requirement becomes $N_x \leq N_C/N_L$, where N_L is the number of spectral channels (that is, wavelengths). Along the vertical direction, to avoid field clipping, the sampling obeys $N_y + N_t - 1 \leq N_R$ because the spatial information and temporal information overlap and occupy the same axis.

Multicolour CUP. To extend CUP's functionality to reproducing colours, we added a spectral separation module in front of the streak camera. As shown in Extended Data Fig. 3a, a dichroic filter ($\geq 98\%$ reflective for 490–554 nm wavelength, 50% reflective and 50% transmissive at 562 nm, and $\geq 90\%$ transmissive for 570–750 nm) is mounted on a mirror at a small tilt angle ($\sim 5^\circ$). The light reflected from this module is divided into two beams according to the wavelength: green light (wavelength < 562 nm) is directly reflected from the dichroic filter, while red light (wavelength

>562 nm) passes through the dichroic filter and bounces from the mirror. Compared with the depth of field of the imaging system, the introduced optical path difference between these two spectral channels is negligible, therefore maintaining the images in focus for both colours.

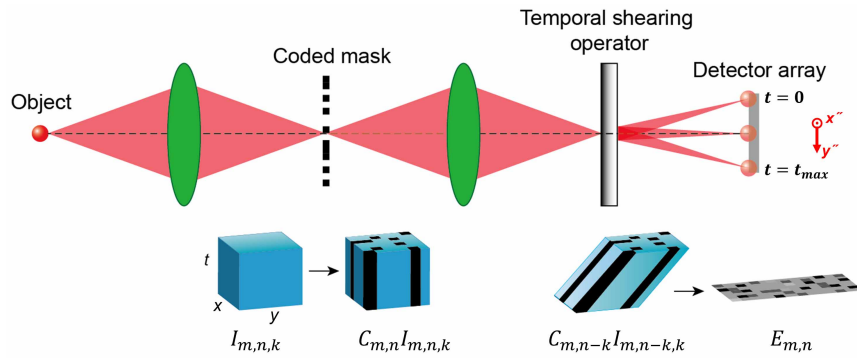
Using the multicolour CUP system, we imaged a pulsed-laser-pumped fluorescence emission process. A fluorophore, Rhodamine 6G, in water solution was excited by a single 7 ps laser pulse at 532 nm. To capture the entire fluorescence decay, we used 50 billion frames per second by setting a shearing velocity of 0.66 mm ns^{-1} on the streak camera. The movie and some representative temporal frames are shown in Supplementary Video 6 and Extended Data Fig. 3b, respectively. In addition, we calculated the time-lapse mean signal intensities within the dashed box in Extended Data Fig. 3b for both the green and red channels (Extended Data Fig. 3c). Based on the measured fluorescence decay, the fluorescence lifetime was found to be 3.8 ns, closely matching a previously reported value³⁰.

In theory, the time delay from the pump laser excitation to the fluorescence emission due to the molecular vibrational relaxation is ~ 6 ps for Rhodamine 6G³¹. However, our results show that the fluorescence starts to decay ~ 180 ps after the pump laser signal reaches its maximum. In the time domain, with 50 billion frames per second sampling, the laser pulse functions as an approximate impulse source while the onset of fluorescence acts as a decaying edge source. Blurring due to the temporal PSF stretches these two signals' maxima apart.

We theoretically simulated this process by using the experimentally measured temporal PSF and the fitted fluorescence decay as the input in Matlab (R2011a). The arrival of the pump laser pulse and the subsequent fluorescence emission are described by a Kronecker delta function (Extended Data Fig. 3d) and an exponentially decaying edge function (Extended Data Fig. 3e), respectively. For the Rhodamine 6G fluorophore, we neglect the molecular vibrational relaxation time and consider the arrival of the pump laser pulse and the onset of fluorescence emission to be

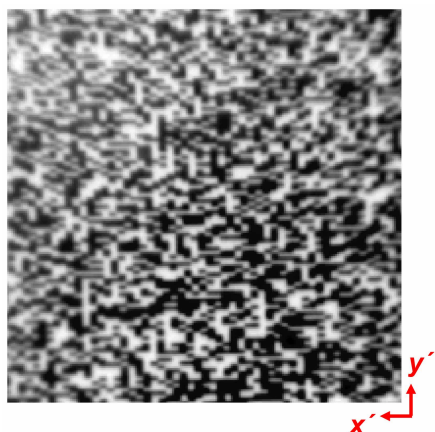
simultaneous. After pump laser excitation, the decay of the normalized fluorescence intensity, $\hat{I}(t)$, is modelled as $\exp(-t/\tau)$, where $\tau = 3.8$ ns. To simulate the temporal-PSF-induced blurring, we convolve an experimentally measured temporal PSF (Extended Data Fig. 3f) with these two event functions shown in Extended Data Fig. 3d and e. The results in Extended Data Fig. 3g indicate that this process introduces an approximate 200 ps time delay between the signal maxima of these two events, which is in good agreement with our experimental observation. Although the full width at half maximum of the main peak in the temporal PSF is only ~ 80 ps, the reconstruction-induced side lobe and shoulder extend over a range of 280 ps, which temporally stretches the signal maxima of these two events apart.

25. Figueiredo, M. A. T., Nowak, R. D. & Wright, S. J. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE Sel. Top. Signal Process.* **1**, 586–597 (2007).
26. Bioucas-Dias, J. M. & Figueiredo, M. A. T. A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Process.* **16**, 2992–3004 (2007).
27. Afonso, M. V., Bioucas-Dias, J. M. & Figueiredo, M. A. T. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. Image Process.* **19**, 2345–2356 (2010).
28. Wright, S. J., Nowak, R. D. & Figueiredo, M. A. T. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Process.* **57**, 2479–2493 (2009).
29. Afonso, M. V., Bioucas-Dias, J. M. & Figueiredo, M. A. T. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Image Process.* **20**, 681–695 (2011).
30. Selinger, K. A., Farnes, J. & Sikkeland, T. Fluorescence lifetime studies of rhodamine 6G in methanol. *J. Phys. Chem.* **81**, 1960–1963 (1977).
31. Rentzepis, P. M., Topp, M. R., Jones, R. P. & Jortner, J. Picosecond emission spectroscopy of homogeneously broadened, electronically excited molecular states. *Phys. Rev. Lett.* **25**, 1742–1745 (1970).

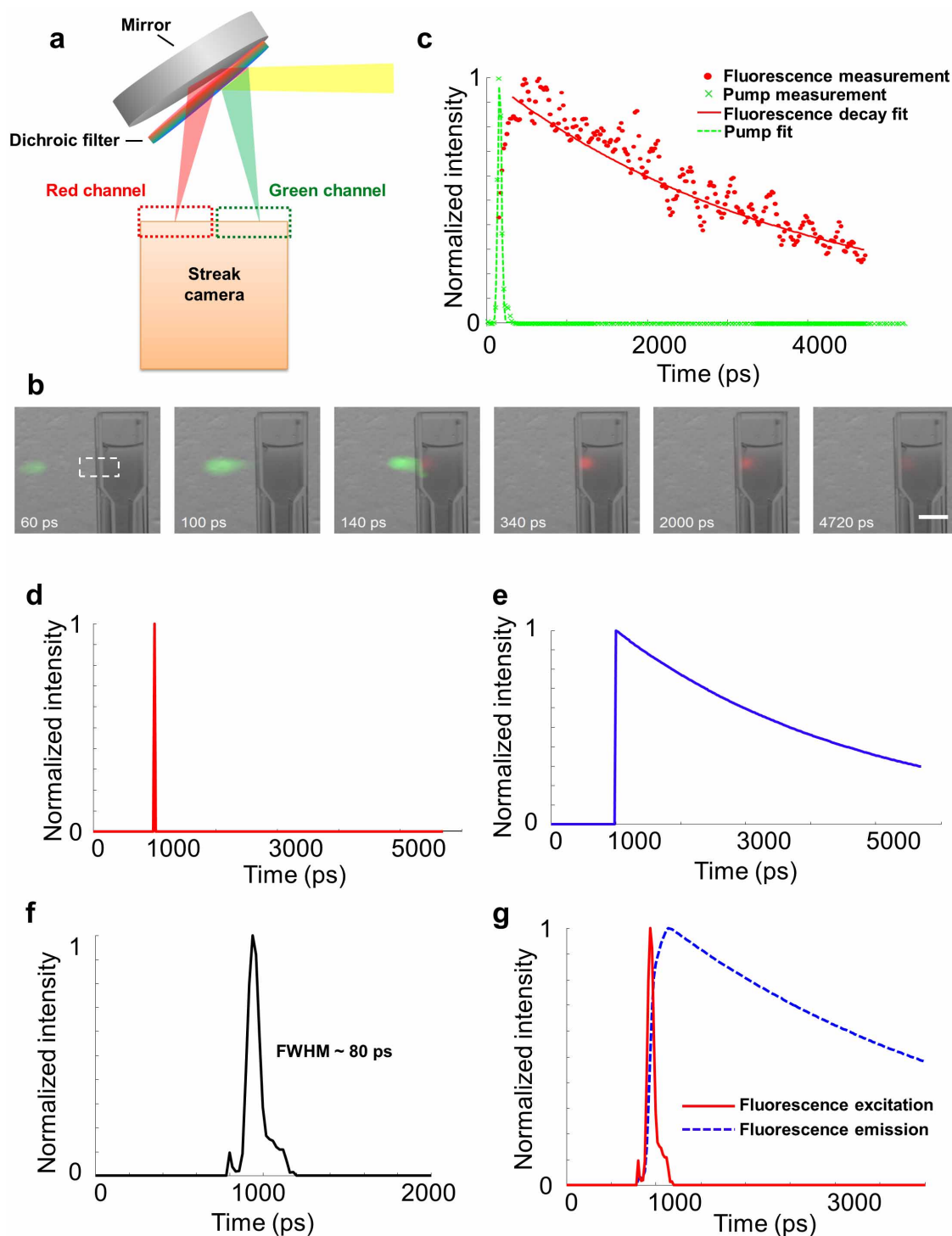


Extended Data Figure 1 | CUP image formation model. x, y , spatial coordinates; t , time; m, n, k , matrix indices; $I_{m,n,k}$, input dynamic scene element; $C_{m,n}$, coded mask matrix element; $C_{m,n-k}I_{m,n-k,k}$, encoded and sheared scene

element; $E_{m,n}$, image element energy measured by a 2D detector array; t_{\max} , maximum recording time. See Methods for details.



Extended Data Figure 2 | A temporally undispersed CCD image of the coded mask, which encodes the uniformly illuminated field with a pseudo-random binary pattern. The position of the mask and definition of the x' - y' plane are shown in Fig. 1.



Extended Data Figure 3 | Multicolour CUP. **a**, Custom-built spectral separation unit. **b**, Representative temporal frames of a pulsed-laser-pumped fluorescence emission process. The pulsed pump laser and fluorescence emission are pseudo-coloured based on their peak emission wavelengths. To explicitly indicate the spatiotemporal pattern of this event, the CUP-reconstructed frames are overlaid with a static background image captured by a monochromatic CCD camera. All temporal frames of this event are provided in Supplementary Video 6. **c**, Time-lapse pump laser and fluorescence emission intensities averaged within the dashed box in **b**. The temporal responses of pump laser excitation and fluorescence decay are fitted to a Gaussian function

and an exponential function, respectively. The recovered fluorescence lifetime of Rhodamine 6G is 3.8 ns. **d**, Event function describing the pulsed laser fluorescence excitation. **e**, Event function describing the fluorescence emission. **f**, Measured temporal PSF, with a full width at half maximum (FWHM) of ~80 ps. Owing to reconstruction artefacts, the PSF has a side lobe and a shoulder extending over a range of 280 ps. **g**, Simulated temporal responses of event functions **d** and **e** after being convolved with the temporal PSF. The maxima of these two time-lapse signals are stretched by 200 ps. Scale bar, 10 mm.

Conductive two-dimensional titanium carbide ‘clay’ with high volumetric capacitance

Michael Ghidui^{1*}, Maria R. Lukatskaya^{1*}, Meng-Qiang Zhao¹, Yury Gogotsi¹ & Michel W. Barsoum¹

Safe and powerful energy storage devices are becoming increasingly important. Charging times of seconds to minutes, with power densities exceeding those of batteries, can in principle be provided by electrochemical capacitors—in particular, pseudocapacitors^{1,2}. Recent research has focused mainly on improving the gravimetric performance of the electrodes of such systems, but for portable electronics and vehicles volume is at a premium³. The best volumetric capacitances of carbon-based electrodes are around 300 farads per cubic centimetre^{4,5}; hydrated ruthenium oxide can reach capacitances of 1,000 to 1,500 farads per cubic centimetre with great cyclability, but only in thin films⁶. Recently, electrodes made of two-dimensional titanium carbide (Ti_3C_2 , a member of the ‘MXene’ family), produced by etching aluminium from titanium aluminium carbide (Ti_3AlC_2 , a ‘MAX’ phase) in concentrated hydrofluoric acid, have been shown to have volumetric capacitances of over 300 farads per cubic centimetre^{7,8}. Here we report a method of producing this material using a solution of lithium fluoride and hydrochloric acid. The resulting hydrophilic material swells in volume when hydrated, and can be shaped like clay and dried into a highly conductive solid or rolled into films tens of micrometres thick. Additive-free films of this titanium carbide ‘clay’ have volumetric capacitances of up to 900 farads per cubic centimetre, with excellent cyclability and rate performances. This capacitance is almost twice that of our previous report⁸, and our synthetic method also offers a much faster route to film production as well as the avoidance of handling hazardous concentrated hydrofluoric acid.

In the search for new electrode materials, two-dimensional solids are of particular interest owing to their large electrochemically active surfaces⁹. For example, activated graphene electrodes have capacitances of 200–350 F cm^{-3} compared to 60–100 F cm^{-3} for activated porous carbons^{10,11}. Yet graphene is limited to the chemistry of carbon, does not tap into metal redox reactions as in ruthenium oxide (RuO_2) (ref. 6), and its conductivity is substantially decreased by the addition of redox-active functional groups¹². MXenes (of the formula $\text{M}_n\text{X}_m\text{T}_x$, where M is a transition metal, X is C and/or N, and T_x denotes surface functionalization) are a relatively young class of two-dimensional solids, produced by the selective etching of the A-group (generally group IIIA and IVA elements) layers from the MAX phases, which comprise a >70-member family of layered, hexagonal early-transition-metal carbides and nitrides¹³. To date, all MXenes have been produced by etching MAX phases in concentrated hydrofluoric acid (HF)^{14–16}.

MXenes have already proved to be promising candidates for electrodes in lithium (Li)-ion batteries^{17,18} and supercapacitors⁸, exhibiting volumetric capacitances that exceed most previously reported materials. However, the path to electrode manufacturing required the handling of concentrated HF and a laborious multi-step procedure. Here we sought a safer route by exploiting the reaction between common, inexpensive hydrochloric acid (HCl) and fluoride salts, leading to dissolution of aluminium and the extraction of two-dimensional carbide layers. Furthermore, given the ability of MXenes to preferentially intercalate cations (post-synthesis)⁸, a related question was whether etching and intercalation might be achieved in a single step, as was observed for etching of

thin Ti_3AlC_2 films with ammonium bifluoride¹⁹. The change in MXene properties upon intercalation and the compositional variability of fluoride salts suggested the possibility of a one-step procedure for the synthesis of many MXenes, with tunable structures and properties.

The MXenes reported in this study were prepared by dissolving LiF in 6 M HCl, followed by the slow addition of Ti_3AlC_2 powders and heating of the mixture at 40 °C for 45 h. After etching, the resulting sediments were washed to remove the reaction products and raise the pH (several cycles of water addition, centrifugation and decanting). The resulting sediment formed a clay-like paste that could be rolled, when wet (Fig. 1a), between water-permeable membranes in a roller mill to produce flexible, free-standing films (Fig. 1c) in a matter of minutes, in contrast to those previously produced by the laborious technique of intercalation, delamination, and filtration¹⁸.

A graphical depiction of the processing is provided in Extended Data Fig. 1. Further, scaling was not limited to the size of the filtration apparatus; films of any dimensions could readily be produced. Additionally, when wet, the ‘clay’ could be moulded and dried to yield various shapes that were highly conductive (Fig. 1d). Diluted, it could also be used as an ink to deposit (print) MXene on various substrates. Like clay, the material could be rehydrated, swelling in volume, and shrinking when dried (Fig. 1b).

Energy-dispersive spectroscopy confirmed that aluminium (Al) was removed, and X-ray diffraction (XRD) revealed the disappearance of Ti_3AlC_2 peaks (traces can be seen in the case of incomplete transformation). Multilayer particles did not show the accordion-like morphology seen in HF-etched MXenes reported to date^{14,20}; rather, particles appeared tightly stacked, presumably as a result of water and/or cationic intercalation (see Extended Data Fig. 2a). Fluorine and oxygen were observed in energy-dispersive spectroscopy; this, coupled with X-ray photoelectron spectroscopy showing evidence of Ti–F and Ti–O bonding, suggests O- and F-containing surface terminations, as has been discussed at length for HF-produced MXenes^{14,21}. The yield of MXene after etching, calculated as described previously¹⁴, is around 100%, which is comparable with the HF-etching method. Our new method thus does not lead to material losses, although an accurate yield determination is difficult owing to the variability of surface groups and amount of intercalated water.

XRD patterns of the etched material, in its air-dried multilayered state, showed a remarkable increase in the intensity and sharpness of the (000l) peaks (Fig. 2a, pink); in some cases the full width at half maximum (FWHM) was as small as 0.188°, as opposed to the broad peaks typical of HF-etched MXene⁷, and more typical of intercalated MXenes¹⁸. Further, compared to a lattice parameter of $c \approx 20 \text{ Å}$ for HF-produced $\text{Ti}_3\text{C}_2\text{T}_x$, the corresponding value in this work was 27–28 Å. XRD patterns of still-hydrated sediment showed shifts to even higher spacings: lattice parameters as high as $c \approx 40 \text{ Å}$ have been measured. These large shifts are suggestive of the presence of water, and possibly cations, between the hydrophilic and negatively charged MXene sheets. From these substantial increases in c and the clay-like properties (see below), it is reasonable to assume that—as in clays^{22,23}—the swelling is due to the intercalation

¹Department of Materials Science and Engineering, and A. J. Drexel Nanomaterials Institute, Drexel University, Philadelphia, Pennsylvania 19104, USA.

*These authors contributed equally to this work.

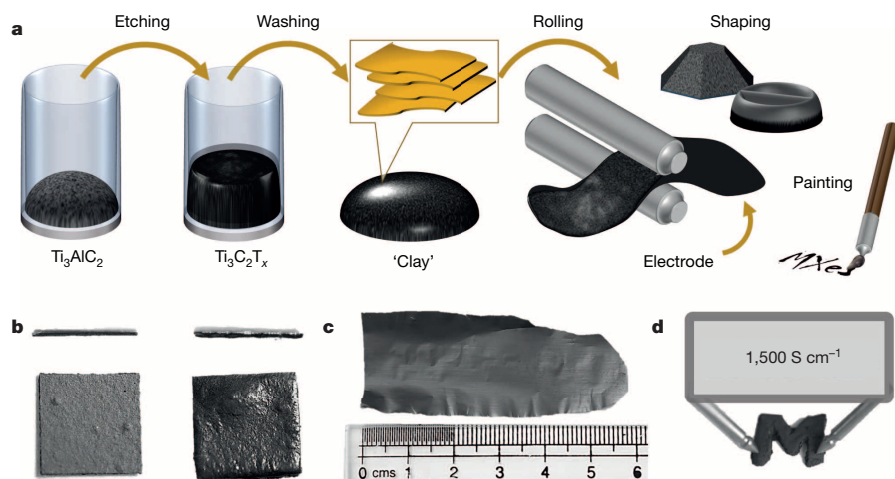


Figure 1 | Schematic of MXene clay synthesis and electrode preparation. **a**, MAX phase is etched in a solution of acid and fluoride salt (step 1), then washed with water to remove reaction products and raise the pH towards neutral (step 2). The resulting sediment behaves like a clay; it can be rolled to produce flexible, freestanding films (step 3), moulded and dried to yield conducting objects of desired shape (step 4), or diluted and painted onto a substrate to yield a conductive coating (step 5). **b**, When dried samples (left, showing cross-section and top view) are hydrated (right) they swell; upon drying, they shrink. **c**, Image of a rolled film. **d**, 'Clay' shaped into the letter M (~ 1 cm) and dried, yielding a conductive solid (labelled with the experimental conductivity of 'clay' rolled to 5 μm thickness). The etched material is referred to as $\text{Ti}_3\text{C}_2\text{T}_x$, where the T denotes surface terminations, such as OH, O and F.

of multiple layers of water and possibly cations between the MXene sheets. Interfacial water has a more structured hydrogen-bonding network than bulk H_2O (ref. 24). The MXene surface, holding a negative electric charge, may act to align the dipoles of water molecules between MXene layers.

When the 'clay' was rolled into freestanding films, XRD patterns again showed strong ordering in the c direction (Fig. 2a, blue). Films, ranging in thicknesses from submicrometre to about 100 μm , were readily produced by this method. The most compelling evidence for particle shearing is the marked intensity decrease of the (110) peak around 61° , indicating a reduction of ordering in the non-basal directions while order in the c direction was maintained (see blue XRD pattern in Fig. 2a and scanning electron microscopy (SEM) image in Extended Data Fig. 2b). Morphologically, the thinner films showed more overall shearing of the multilayer particles when viewed in cross-section (Fig. 2e, f) and exhibited substantial flexibility, even when allowed to dry thoroughly (inset to Fig. 2e). The contact angle of water on the rolled MXene film was measured as 21.5° , confirming its hydrophilic nature (Extended Data Fig. 3). Attempts to hydrate and roll HF-produced MXene were unsuccessful; we propose that the intercalated water acts as a lubricant that allows facile shearing.

The c parameter expansion also resulted in the weakening of interactions between the MXene layers, as evidenced by the easy delamination of multilayered particles by sonication, as is done for van der Waals solids⁹. In our previous work, typical sonication times for delamination (after post-synthesis intercalation with dimethyl sulphoxide) were of the order of 4 h (ref. 18). Here, sonication times of the order of 30–60 min resulted in stable suspensions with concentrations as high as 2 g per litre, higher than observed previously. Remarkably, the yield from multilayer to dispersed flakes was about 45% by mass. Freestanding films were also readily fabricated by filtering these suspensions, as reported previously⁸.

The fact that the $\text{LiF} + \text{HCl}$ etchant was much milder than HF resulted in flakes with larger lateral dimensions (Fig. 2b) that did not contain the nanometre-size defects frequently observed in HF-etched samples²⁵. Transmission electron microscopy (TEM) analysis showed that, of 321 flakes analysed, over 70% had dimensions of 0.5–1.5 μm (Extended Data Fig. 4a, b). Single layers about 10 Å thick were imaged using TEM (Fig. 2c, d), confirming that the material is indeed two-dimensional. Analysis of 332 flakes suggested that roughly 70% of the flakes were 1–2 layers thick (Extended Data Fig. 4c–f). We note that, since the restacking or folding of flakes can lead to higher apparent thicknesses (Extended Data Fig. 5), the 70% estimate is conservative. Thus, using this method,

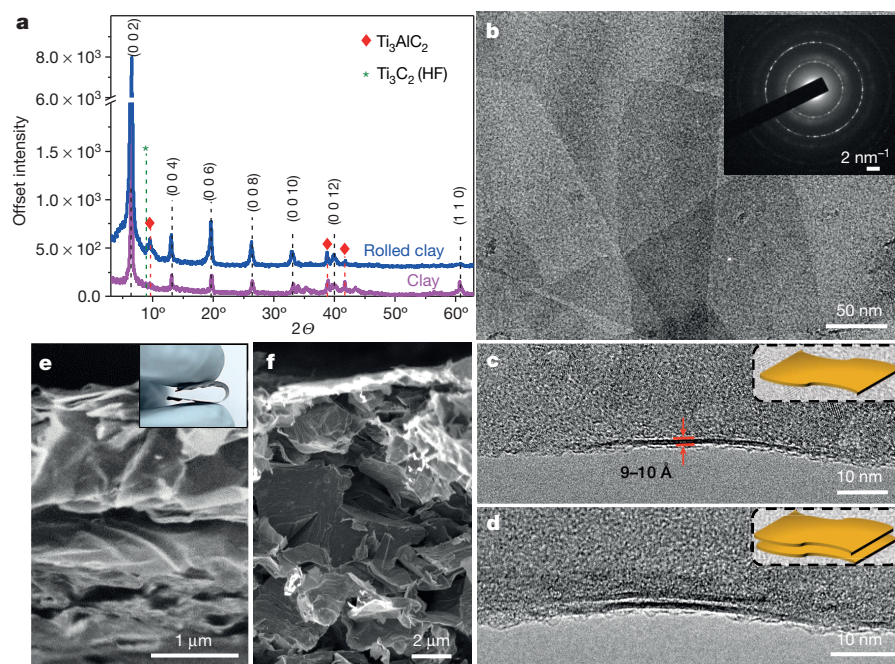


Figure 2 | Structural characterization of MXene. **a**, XRD patterns of samples produced by etching in $\text{LiF} + \text{HCl}$ solution. The pink trace is for multilayer $\text{Ti}_3\text{C}_2\text{T}_x$, showing a sharp, intense peak (0002) and higher-order (000 n) peaks, corresponding to a c lattice parameter of 28 Å and high order in the c direction. The blue trace is for the same sample after rolling into an approximately 40- μm -thick film; c -direction peaks are preserved, but the prominent (110) peak is no longer observed, showing substantial reduction of order in non-basal directions. In both cases, traces of Ti_3AlC_2 are still present (red diamonds). The MXene (0002) peak is at a much lower angle than that typical of MXene produced by HF (green star). **b**, TEM image of several flakes, showing lateral sizes up to a few hundred nanometres. Few defective areas are present. The inset shows the overall selected area electron diffraction pattern. **c**, **d**, TEM images of single- and double-layer flakes, respectively. Insets show sketches of these layers. **e**, SEM image of a fracture surface of a ~ 4 - μm -thick film produced by rolling, showing shearing of layers; the flexibility of the film is demonstrated in the inset. **f**, Fracture surface of a thicker rolled film (~ 30 μm), showing poorer overall alignment of flakes in the interior of the film.

large fractions of single-layered MXene flakes with high yields, large lateral sizes, and good quality can be readily produced. The flake lateral sizes reported here are larger than those reported for HF-etched Ti_3AlC_2 (ref. 18); the milder delamination conditions may be partially responsible for this.

Previously we have shown that MXene ‘paper’—made by filtration of solutions containing delaminated $\text{Ti}_3\text{C}_2\text{T}_x$ flakes—exhibited volumetric capacitances of $\sim 350 \text{ F cm}^{-3}$ at 20 mV s^{-1} (and 450 F cm^{-3} at 2 mV s^{-1}) in potassium hydroxide (KOH) electrolyte⁸. For comparison, we characterized the electrochemical performances of rolled, freestanding $\text{Ti}_3\text{C}_2\text{T}_x$ films in 1 M sulphuric acid (H_2SO_4). The advantages of acidic electrolytes include not only their excellent conductivities but also that protons, being the smallest cations, are known to allow for surface redox reactions in transition-metal oxide electrodes, such as RuO_2 , MnO_2 and some others, and may contribute to the capacitance via fast surface redox^{1,26}.

At a scan rate of 2 mV s^{-1} , capacitance values reached 900 F cm^{-3} (Fig. 3a) and a good rate handling ability was observed (Fig. 3b). The results—summarized and compared with previous work⁸ in Fig. 3b—clearly show that rolled $\text{Ti}_3\text{C}_2\text{T}_x$ clay electrodes show outstanding capacitive performance, not only volumetrically but gravimetrically as well, achieving 245 F g^{-1} at 2 mV s^{-1} . This can be ascribed to the smaller size of H^+ compared to other intercalating cations, surface redox processes, and improved accessibility of interlayer spacing in $\text{LiF} + \text{HCl}$ -etched MXene owing to pre-intercalated water, compared to the previously studied HF-etched samples. It is worth noting that a similar positive role of structural water for capacitive performance in acidic electrolytes was observed for hydrated ruthenium oxide²⁷. The electrodes showed no measurable capacitance losses even after 10,000 cycles (Fig. 3c). Coulombic efficiency is close to 100% (inset in Fig. 3c), confirming that the outstanding performance is not due to parasitic reactions.

To quantify the capacitive and diffusion limited contributions to the total capacitances, we used the approach of ref. 28. The results of

this analysis—summarized in Fig. 3d—show that, at scan rates below 20 mV s^{-1} , there is a noticeable, yet not prevailing, contribution of diffusion-limited processes to the total capacitance. At scan rates of 20 mV s^{-1} and higher, the response is not diffusion-controlled but is rather due to surface capacitive effects, whether electrostatic or pseudocapacitive. This observation is in agreement with the conjecture of Levi *et al.*²⁹ about the presence of shallow- and deep-trap sites in MXene structures. Further, if there are also redox contributions from changes in the oxidation states of surface Ti atoms²⁸, the redox processes are not diffusion-limited, and thus represent ‘intrinsic’ capacitive behaviour³⁰.

When the electrochemical responses of three rolled clay electrodes ($5 \mu\text{m}$ thick, $30 \mu\text{m}$ thick and $75 \mu\text{m}$ thick) were compared (Fig. 3e, f), not surprisingly, the volumetric capacitances decreased with increased thickness. These thickness-dependent differences can be partially traced to the electrode morphologies. As noted above, electrodes thinner than $10 \mu\text{m}$ showed good flake alignment (Fig. 2e) with typical densities of $3.6\text{--}3.8 \text{ g cm}^{-3}$. At $2.2\text{--}2.8 \text{ g cm}^{-3}$, the densities of the thicker ($15 \mu\text{m}$ and larger) rolled electrodes were lower, which is a reflection of the fact that their core seemed to be more open (Fig. 2f). And while the lower densities led to lower volumetric capacitances, their more open structure ensured accessibility to ions and thus similar rate performances as their thinner counterparts (Fig. 3e, f). The lower densities also ensured that the drop in gravimetric capacitances with thickness (see Extended Data Fig. 6) was not substantial. A summary of key mass- and volume-normalized capacitance values as a function of electrode thickness is provided in Extended Data Table 1. Although the voltage window used for testing is relatively narrow, it can be expanded by conducting tests in other types of electrolytes, such as neutral aqueous and organic electrolytes, or using MXenes as negative electrodes in asymmetric cell configurations.

The good capacitive rate performance of the $75\text{-}\mu\text{m}$ -thick electrodes (Fig. 3e) is noteworthy, however, and demonstrates scalability and huge promise of MXenes for application as negative electrodes of hybrid

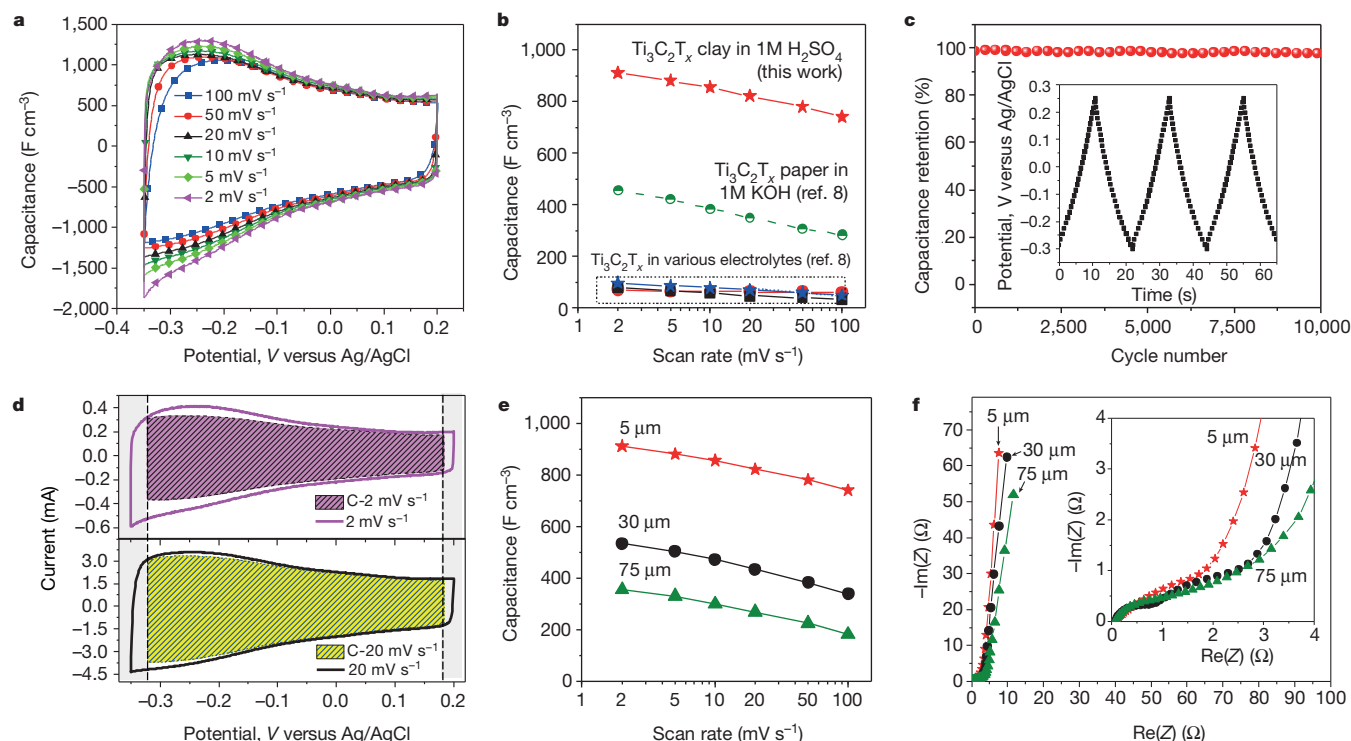


Figure 3 | Electrochemical performance of rolled, free-standing electrodes. **a**, Cyclic voltammetry profiles at different scan rates for a $5\text{-}\mu\text{m}$ -thick electrode in 1 M H_2SO_4 . **b**, Comparison of rate performances reported in this work and previously for HF-produced MXene⁸. **c**, Capacitance retention test of a $5\text{-}\mu\text{m}$ -thick rolled electrode in 1 M H_2SO_4 . Inset shows galvanostatic cycling data collected at 10 A g^{-1} . **d**, Cyclic voltammetry profiles collected at 2 mV s^{-1}

and 20 mV s^{-1} with hatched portions of the contributions of the processes not limited by diffusion, that is, capacitive (‘C-’); vertical lines limit the cyclic voltammetry area used in calculations. **e**, **f**, Rate performance (**e**) and electrochemical impedance spectroscopy data (**f**) of $5\text{-}\mu\text{m}$ -thick (red stars), $30\text{-}\mu\text{m}$ -thick (black circles) and $75\text{-}\mu\text{m}$ -thick (olive triangles) rolled electrodes. The inset in **f** shows the magnified high-frequency region.

large-scale energy storage devices. Electrodes of that thickness cannot be produced by filtration and the MXene clay-like characteristics add important versatility to electrode manufacturing, allowing films of the required thicknesses to be rolled. Note that the capacitance values reported herein are still preliminary. As better understanding of how the films' morphologies affect their capacitances is gained, enhancements in the latter should ensue.

In terms of versatility, the LiF + HCl solution was also capable of etching other MAX phases, for example, Nb₂AlC and Ti₂AlC. In the case of Ti₂AlC, we delaminated the multilayer powders in a similar fashion to Ti₃C₂T_x to produce suspensions of Ti₂CT_x flakes, as well as Ti₂CT_x 'paper', which had not been previously reported. These considerations hint at the potential of this new etching method for the synthesis of other MXenes, which will be explored in future studies.

This method of MXene production was successful to varying degrees for other fluoride salts, such as NaF, KF, CsF, tetrabutylammonium fluoride, and CaF₂ in HCl, all of which showed similar etching behaviour. When H₂SO₄ was used instead of HCl, MXenes were still obtained. We note here that these systems are options and merit further study; the ability to fine tune the reaction based on reagents used will indubitably lead to potentially useful variations in compositions and properties, especially since it is reasonable to assume that different acids and salts should modify the surface chemistries and pre-intercalate different ions.

In summary, a new high-yield method for MXene synthesis that is safer, easier, and provides a faster route to delaminated flakes has been detailed. This method yields a clay-like material (for a discussion of the effect of experimental conditions on properties, see Methods), which can be shaped to give conductive solids of desired forms, or rolled into thin sheets, for a host of applications. When the rolled films were used as supercapacitor electrodes in a H₂SO₄ electrolyte, the performances were extraordinary, with volumetric capacitances up to 900 F cm⁻³ or 245 F g⁻¹. When it is further appreciated that these numbers are 'first-generation' numbers that will no doubt increase as we better understand the underlying processes and modify the material structure and chemistry, the potential of these non-oxide two-dimensional materials to push electrochemical energy storage to new heights is clear.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 August; accepted 13 October 2014.

Published online 26 November 2014.

- Simon, P. & Gogotsi, Y. Materials for electrochemical capacitors. *Nature Mater.* **7**, 845–854 (2008).
- Augustyn, V. *et al.* High-rate electrochemical energy storage through Li⁺ intercalation pseudocapacitance. *Nature Mater.* **12**, 518–522 (2013).
- Gogotsi, Y. & Simon, P. True performance metrics in electrochemical energy storage. *Science* **334**, 917–918 (2011).
- Murali, S. *et al.* Volumetric capacitance of compressed activated microwave-expanded graphite oxide (a-MEGO) electrodes. *Nano Energy* **2**, 764–768 (2013).
- Yang, X., Cheng, C., Wang, Y., Qiu, L. & Li, D. Liquid-mediated dense integration of graphene materials for compact capacitive energy storage. *Science* **341**, 534–537 (2013).
- Zheng, J. P., Cygan, P. J. & Jow, T. R. Hydrous ruthenium oxide as an electrode material for electrochemical capacitors. *J. Electrochem. Soc.* **142**, 2699–2703 (1995).
- Naguib, M. *et al.* Two-dimensional nanocrystals produced by exfoliation of Ti₃AlC₂. *Adv. Mater.* **23**, 4248–4253 (2011).
- Lukatskaya, M. R. *et al.* Cation intercalation and high volumetric capacitance of two-dimensional titanium carbide. *Science* **341**, 1502–1505 (2013).
- Nicolosi, V., Chhowalla, M., Kanatzidis, M. G., Strano, M. S. & Coleman, J. N. Liquid exfoliation of layered materials. *Science* **340**, 6139 (2013).

- Ghaffari, M. *et al.* High-volumetric performance aligned nano-porous microwave exfoliated graphite oxide-based electrochemical capacitors. *Adv. Mater.* **25**, 4879–4885 (2013).
- Tao, Y. *et al.* Towards ultrahigh volumetric capacitance: graphene derived highly dense but porous carbons for supercapacitors. *Sci. Rep.* **3**, 2975 (2013).
- Jung, I., Dikin, D. A., Piner, R. D. & Ruoff, R. S. Tunable electrical conductivity of individual graphene oxide sheets reduced at 'low' temperatures. *Nano Lett.* **8**, 4283–4287 (2008).
- Barsoum, M. W. *MAX Phases: Properties of Machinable Ternary Carbides and Nitrides* (John Wiley & Sons, 2013).
- Naguib, M., Mochalin, V. N., Barsoum, M. W. & Gogotsi, Y. MXenes: a new family of two-dimensional materials. *Adv. Mater.* **26**, 982 (2014).
- Xie, X. *et al.* Surface Al leached Ti₃AlC₂ substituting carbon for catalyst support served in a harsh corrosive electrochemical system. *Nanoscale* **6**, 11035–11040 (2014).
- Peng, Q. *et al.* Unique lead adsorption behavior of activated hydroxyl group in two-dimensional titanium carbide. *J. Am. Chem. Soc.* **136**, 4113–4116 (2014).
- Tang, Q., Zhou, Z. & Shen, P. Are MXenes promising anode materials for Li ion batteries? computational studies on electronic properties and Li storage capability of Ti₃C₂ and Ti₃C₂X₂ (X = F, OH) monolayer. *J. Am. Chem. Soc.* **134**, 16909–16916 (2012).
- Mashtalir, O. *et al.* Intercalation and delamination of layered carbides and carbonitrides. *Nature Commun.* **4**, 1716 (2013).
- Halim, J. *et al.* Transparent conductive two-dimensional titanium carbide epitaxial thin films. *Chem. Mater.* **26**, 2374–2381 (2014).
- Chang, F., Li, C., Yang, J., Tang, H. & Xue, M. Synthesis of a new graphene-like transition metal carbide by de-intercalating Ti₃AlC₂. *Mater. Lett.* **109**, 295–298 (2013).
- Enyashin, A. N. & Ivanovskii, A. L. Two-dimensional titanium carbonitrides and their hydroxylated derivatives: structural, electronic properties and stability of MXenes Ti₃C₂–_xN_x(OH)₂ from DFTB calculations. *J. Solid State Chem.* **207**, 42–48 (2013).
- Madsen, F. T. & Müller-Vonmoos, M. The swelling behaviour of clays. *Appl. Clay Sci.* **4**, 143–156 (1989).
- Hensen, E. J. & Smit, B. Why clays swell. *J. Phys. Chem. B* **106**, 12664–12667 (2002).
- Lis, D., Backus, E. H. G., Hunger, J., Parekh, S. H. & Bonn, M. Liquid flow along a solid surface reversibly alters interfacial chemistry. *Science* **344**, 1138–1142 (2014).
- Mashtalir, O., Naguib, M., Dyatkin, B., Gogotsi, Y. & Barsoum, M. W. Kinetics of aluminum extraction from Ti₃AlC₂ in hydrofluoric acid. *Mater. Chem. Phys.* **139**, 147–152 (2013).
- Conway, B. Electrochemical capacitors based on pseudocapacitance. In *Electrochemical Supercapacitors: Scientific Fundamentals and Technological Applications* (Kluwer Academic/Plenum, 1999).
- Dmowski, W., Egami, T., Swider-Lyons, K. E., Love, C. T. & Rolison, D. R. Local atomic structure and conduction mechanism of nanocrystalline hydrous RuO₂ from X-ray scattering. *J. Phys. Chem. B* **106**, 12677–12683 (2002).
- Wang, J., Polleux, J., Lim, J. & Dunn, B. Pseudocapacitive contributions to electrochemical energy storage in TiO₂ (anatase) nanoparticles. *J. Phys. Chem. C* **111**, 14925–14931 (2007).
- Levi, M. D. *et al.* Solving the capacitive paradox of 2D MXene by electrochemical quartz-crystal admittance and in situ electronic conductance measurements. *Adv. Energy Mater.* <http://dx.doi.org/10.1002/aenm.201400815> (2014).
- Simon, P., Gogotsi, Y. & Dunn, B. Where do batteries end and supercapacitors begin? *Science* **343**, 1210–1211 (2014).

Acknowledgements We thank O. Mashtalir and Z. Ling for help with material characterization. This work was supported by the US National Science Foundation under grant number DMR-1310245. Electrochemical research was supported by the Fluid Interface Reactions, Structures and Transport (FIRST) Center, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, and Office of Basic Energy Sciences. XRD, X-ray photoelectron spectroscopy, SEM and TEM investigations were performed at the Centralized Research Facilities at Drexel University.

Author Contributions M.G. conducted material synthesis and XRD analysis. M.R.L. performed electrochemical measurements and SEM analysis. M.-Q.Z. performed TEM analysis. M.W.B. and Y.G. planned and supervised the research. M.R.L., M.G., M.W.B. and Y.G. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B. (barsoumw@drexel.edu) or Y.G. (gogotsi@drexel.edu).

METHODS

Synthesis of Ti_3AlC_2 . The MAX phase used as precursor for MXene synthesis herein— Ti_3AlC_2 —was prepared by mixing commercial Ti_2AlC powders (Kanthal, Sweden) with TiC in a 1:1 molar ratio (after adjusting for the ~12 wt% Ti_3AlC_2 already present in the commercial powder), followed by ball milling for 18 h. The mixture was then heated at 5°C min^{-1} , under flowing argon (Ar) in a tube furnace for 2 h at $1,350^\circ\text{C}$. The resulting lightly sintered brick was ground with a TiN-coated milling bit and sieved through a 400 mesh sieve producing powder with particle size less than $38\text{ }\mu\text{m}$.

Synthesis of $\text{Ti}_3\text{C}_2\text{T}_x$ MXene. Concentrated HCl (Fisher, technical grade), was added to distilled water to prepare a 6 M solution (30 ml total). 1.98 g (5 molar equivalents) of LiF (Alfa Aesar, 98+%) was added to this solution. The mixture was stirred for 5 min with a magnetic Teflon stir bar to dissolve the salt.

Three grams of Ti_3AlC_2 powders were carefully added over the course of 10 min to avoid initial overheating of the solution as a result of the exothermic nature of the reactions. The reaction mixture was then held at 40°C for 45 h, after which the mixture was washed through ~5 cycles of distilled water addition, centrifugation ($3,500\text{ r.p.m.} \times 5\text{ min}$ for each cycle), and decanting, until the supernatant reached a pH of approximately 6. The final product, with a small amount of water, was filtered on cellulose nitrate ($0.22\text{ }\mu\text{m}$ pore size). At this stage, the filtrate exhibited clay-like properties and could be directly processed into films by rolling.

Preparation of $\text{Ti}_3\text{C}_2\text{T}_x$ 'paper'. The $\text{Ti}_3\text{C}_2\text{T}_x$ flakes were dispersed in distilled water (2 g MXene per 0.5 litre of water), deaerated with Ar, followed by sonication for 1 h. The mixture was then centrifuged for 1 h at $3,500\text{ r.p.m.}$, and the supernatant, which was dark green in colour, was collected. This dispersion was filtered using a membrane (3501 Coated PP, Celgard, USA) to yield flexible, freestanding $\text{Ti}_3\text{C}_2\text{T}_x$ paper. The weight percentage of MXene delaminated into stable suspension in this case was around 45 wt%.

$\text{Ti}_3\text{C}_2\text{T}_x$ clay electrodes. Preparation of the clay electrodes is depicted step-by-step in Extended Data Fig. 1. The dried and crushed $\text{Ti}_3\text{C}_2\text{T}_x$ powder is hydrated to the consistency of a thick paste, roughly two parts powder to one part water (Extended Data Fig. 1a–c), which turns it into a plastic, clay-like state, that can be formed and moulded. The 'clay' is then rolled using a roller mill with water-permeable Celgard sheets (Extended Data Fig. 1d) on either side, resulting in the formation of a free-standing film (Extended Data Fig. 1e), which was readily lifted off the membrane upon drying (Extended Data Fig. 1f).

Activated carbon electrodes. The activated carbon electrodes were prepared by mechanical processing of a pre-mixed slurry, containing ethanol (190 proof, Decon Laboratories), YP-50 activated carbon powder (Kuraray, Japan), and polytetrafluoroethylene (PTFE) binder (60 wt% in H_2O , Sigma Aldrich). The resulting composition of the activated carbon electrodes was 95 wt% activated carbon and 5 wt% PTFE. They had thicknesses that varied between $100\text{ }\mu\text{m}$ and $150\text{ }\mu\text{m}$; the mass densities per unit area were in the $10\text{--}25\text{ mg cm}^{-2}$ range.

Electrochemical setup. All electrochemical measurements were performed in three-electrode Swagelok cells, in which MXene served as the working electrode, over-capacitive activated carbon films were used as the counter electrode, and Ag/AgCl in 1 M KCl was the reference electrode. Two layers of the Celgard membranes were used as separators. The electrolyte was 1 M H_2SO_4 (Alfa Aesar, American Chemical Society grade).

Electrochemical measurements. Cyclic voltammetry, electrochemical impedance spectroscopy, and galvanostatic cycling were performed using a VMP3 potentiostat (Biologic, France). Cyclic voltammetry was performed using scan rates that ranged

from 1 mV s^{-1} to 100 mV s^{-1} . Electrochemical impedance spectroscopy was performed at open circuit potential, with a 10-mV amplitude, and frequencies that ranged from 10 mHz to 200 kHz. Galvanostatic cycling was performed at 1 A g^{-1} and 10 A g^{-1} between the potential limits of -0.3 V to 0.25 V versus Ag/AgCl. Capacitance data reported in the article were calculated from the slope of the discharge curve.

Characterization of structure and properties. XRD patterns were recorded with a powder diffractometer (Rigaku SmartLab) using Cu K_α radiation ($\lambda = 1.54\text{ }\text{\AA}$) with 0.2° 2θ steps and 0.5 s dwelling time.

Scanning electron microscopy was performed on a Zeiss Supra 50VP (Carl Zeiss SMT AG, Oberkochen, Germany) equipped with an energy-dispersive spectroscope (Oxford EDS, with INCA software). Most energy-dispersive spectroscope scans were obtained at low magnification ($100\times$ to $200\times$) at random points of powdered samples.

Transmission electron microscopy of the MXene flakes was performed on a JEM-2100 (JEOL, Japan) using an accelerating voltage of 200 kV. The TEM samples were prepared by dropping two drops of diluted colloidal solution of MXene flakes onto a copper grid and drying in air. The flake size and number of layers per flake distributions were obtained through statistical analysis of more than 300 MXene flakes in the TEM images.

Resistivity measurements were performed with a 4-point probe (ResTest v1, Jandel Engineering, UK). Measured resistivity was automatically multiplied by the proper thickness correction factor given by the Jandel software.

Temperature and time in the MXene synthesis. We found that reaction conditions of 35°C for 24 h rather than 40°C for 45 h produced a material with persistent MAX-phase peaks in the XRD patterns, and higher Al content revealed by energy-dispersive spectroscopy, but that gave reliable high yields of delaminated flakes upon sonication. The Ti_3AlC_2 etched at higher temperatures showed lower Al content but did not always readily delaminate and disperse by sonication.

Capacitance calculations. The volumetric capacitance determined from the cyclic voltammetry data is given by:

$$C = \frac{1}{\Delta V} \int \frac{j dV}{s} \quad (1)$$

and the volumetric capacitance determined from the galvanostatic charge/discharge data is given by:

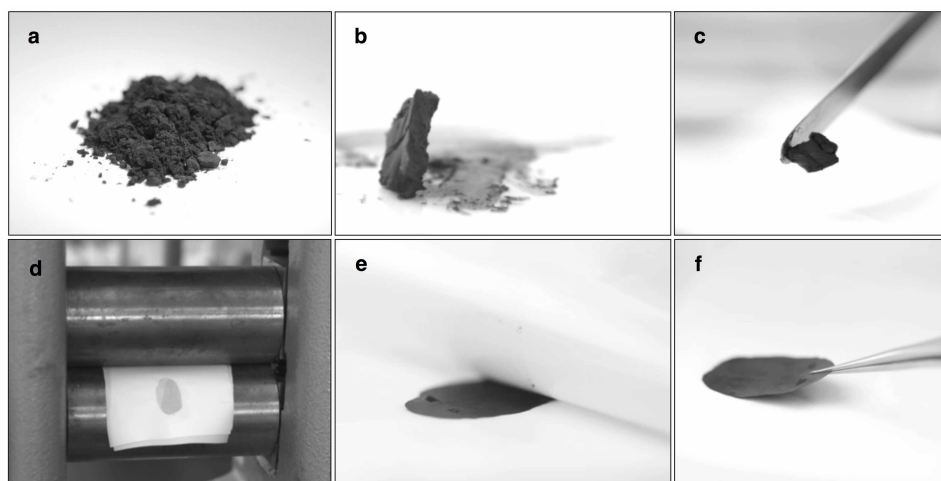
$$C = (jt)/V \quad (2)$$

where C is the normalized capacitance (in units of F cm^{-3}), j is the current density (in A cm^{-2}), s is the rate (in V s^{-1}), V is the voltage (in V), ΔV is the voltage window (in V) and t is time (in s). Calculations of the gravimetric capacitance (in F g^{-1}) were performed using current density per electrode weight (in A g^{-1}).

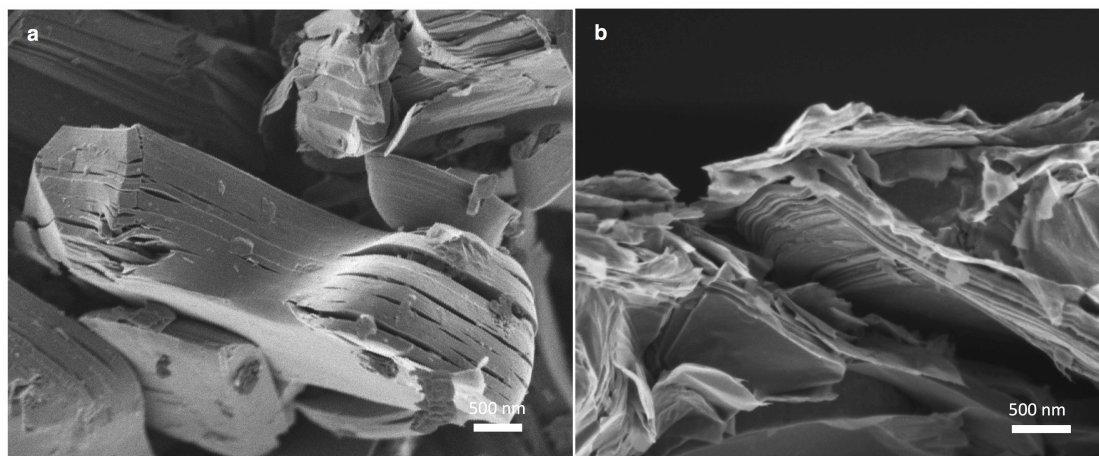
Analysis of the limiting processes for charge storage. To quantify the diffusion-limited contribution to capacitance, the relationship between the current $i(V)$ (at a given voltage V , in units of mA) and scan rate v (in units of V s^{-1}), was assumed to be²⁷:

$$i(V) = k_1 v + k_2 v^{0.5}$$

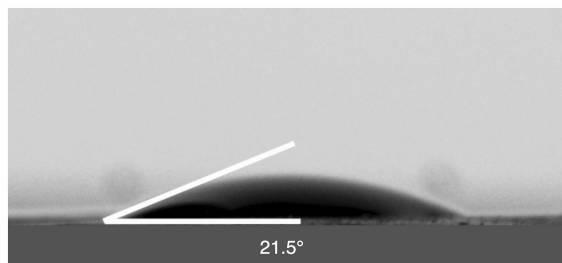
where k_1 and k_2 are constants. For the cyclic voltammetry experiments, at scan rates from 1 mV s^{-1} to 20 mV s^{-1} , current values were extracted, and $i/v^{0.5}$ versus $v^{0.5}$ was plotted at each voltage and linear fitting was performed: $i(V)/v^{0.5} = k_1 v^{0.5} + k_2$. The slope k_1 , for each voltage, describes the contributions of the non-diffusion controlled processes to the overall process.



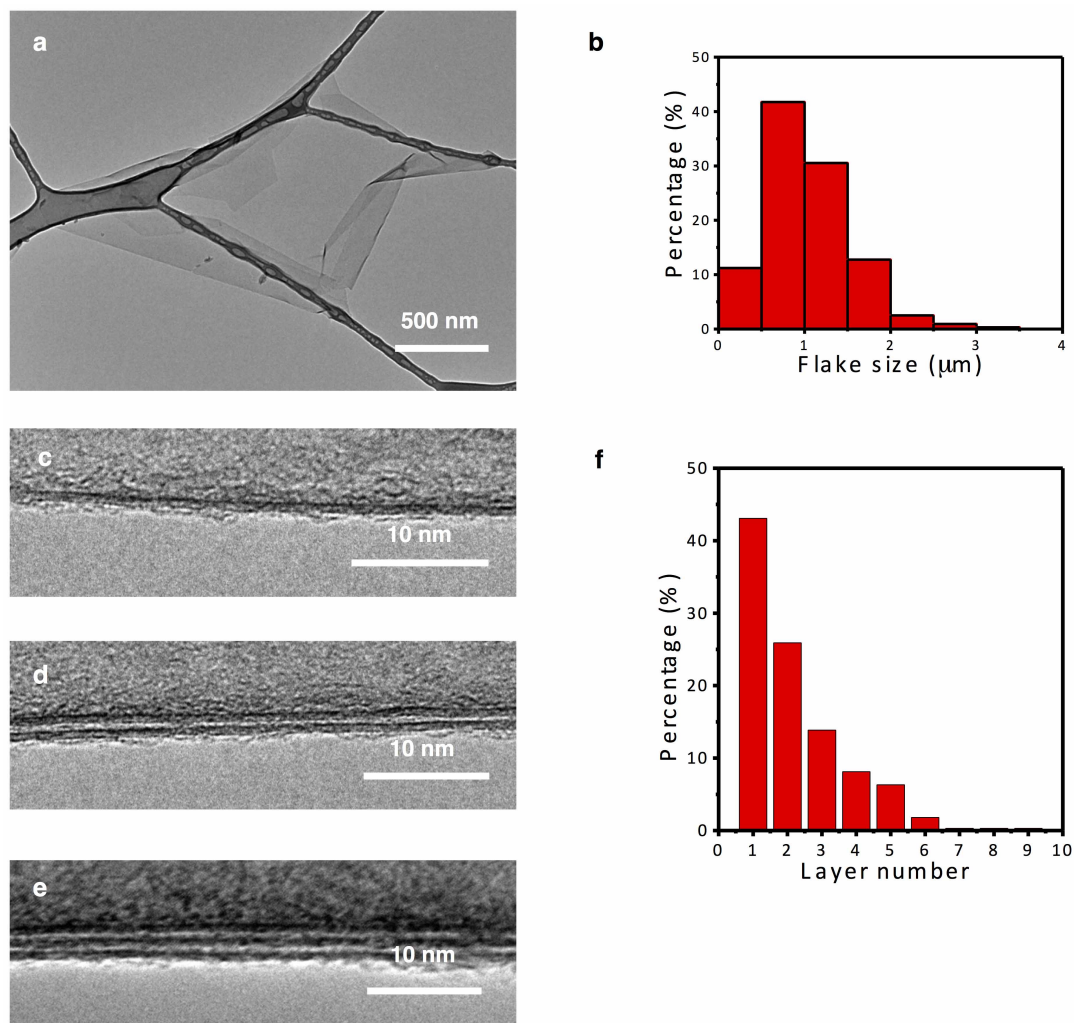
Extended Data Figure 1 | Processing of MXene clay. **a**, Dried and crushed powder. **b**, **c**, Hydrated clay is plastic and can be readily formed and moulded. **d**, Demonstration of films produced in the roller mill. **e**, **f**, Rolled freestanding film being lifted off Celgard membranes.



Extended Data Figure 2 | SEM images. **a**, Multilayer MXene particle. **b**, Cross-section of rolled Ti₃C₂ film, showing shearing that is most probably responsible for the loss of the 60° angle peak in the XRD pattern.



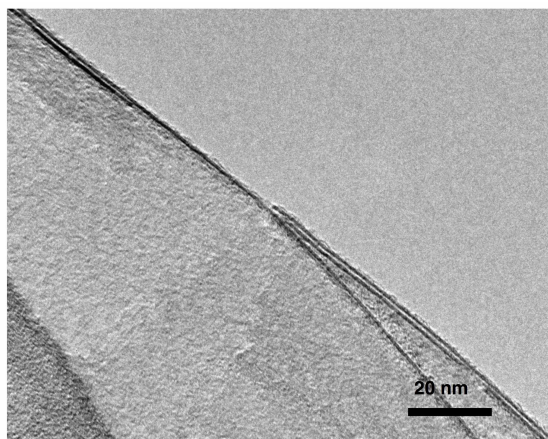
Extended Data Figure 3 | Contact angle. Digital image showing contact angle of a water droplet on rolled MXene film, indicating its hydrophilic surface.



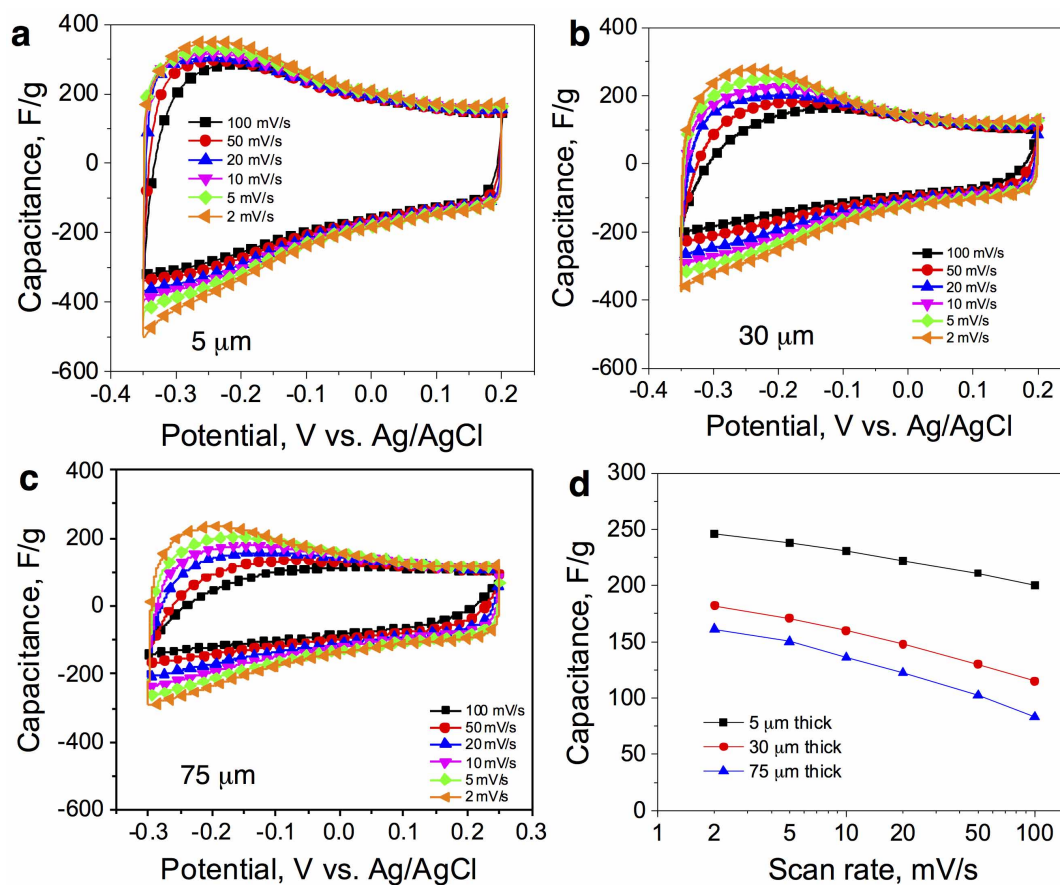
Extended Data Figure 4 | TEM characterization of dispersed $\text{Ti}_3\text{C}_2\text{T}_x$ flakes.

a, Representative TEM image showing the morphology and size of a large single-layer $\text{Ti}_3\text{C}_2\text{T}_x$ flake. Note folding on all sides of this large flake. **b**, The lateral size distribution of the dispersed $\text{Ti}_3\text{C}_2\text{T}_x$ flakes. **c–e**, Representative TEM images showing single-layer (**c**), double-layer (**d**) and triple-layer (**e**)

flakes. **f**, Statistical analysis of layer number distribution of dispersed $\text{Ti}_3\text{C}_2\text{T}_x$ flakes. Note that the fractions of double- and few-layer flakes are overestimated owing to inevitable restacking and edge folding of single-layer flakes during preparation of samples for TEM analysis. Edge folding is clearly seen in **a**. An example of restacking is shown in Extended Data Fig. 5.



Extended Data Figure 5 | TEM image showing the restacking of single- or double-layer MXene flakes into few-layer MXene.



Extended Data Figure 6 | Gravimetrically normalized capacitance. Cyclic voltammetry profiles at different scan rates for 5- μm -thick (a), 30- μm -thick (b) and 75- μm -thick (c) electrodes in 1 M H_2SO_4 . d, Gravimetric

rate performances of rolled electrodes, 5 μm thick (black squares), 30 μm thick (red circles) and 75 μm thick (blue triangles).

Extended Data Table 1 | Effect of film thickness and scan rate on mass- and volume-normalized capacitance values

Scan rate, mV/s	5 μm electrode		30 μm electrode		75 μm electrode	
	F/ cm ³	F/g	F/ cm ³	F/g	F/ cm ³	F/g
2	910	246	534	182	355	161
5	881	238	504	171	330	150
10	855	231	472	160	300	136
20	821	222	434	148	269	122
50	781	210	383	130	225	102
100	740	200	339	115	183	83

Eastern Pacific tropical cyclones intensified by El Niño delivery of subsurface ocean heat

F.-F. Jin^{1,3}, J. Boucharel¹ & I.-I. Lin²

The El Niño Southern Oscillation (ENSO) creates strong variations in sea surface temperature in the eastern equatorial Pacific, leading to major climatic and societal impacts^{1,2}. In particular, ENSO influences the yearly variations of tropical cyclone (TC) activities in both the Pacific and Atlantic basins through atmospheric dynamical factors such as vertical wind shear and stability^{3–6}. Until recently, however, the direct ocean thermal control of ENSO on TCs has not been taken into consideration because of an apparent mismatch in both timing and location: ENSO peaks in winter and its surface warming occurs mostly along the Equator, a region without TC activity. Here we show that El Niño—the warm phase of an ENSO cycle—effectively discharges heat into the eastern North Pacific basin two to three seasons after its wintertime peak, leading to intensified TCs. This basin is characterized by abundant TC activity and is the second most active TC region in the world^{5–7}. As a result of the time involved in ocean transport, El Niño's equatorial subsurface 'heat reservoir', built up in boreal winter, appears in the eastern North Pacific several months later during peak TC season (boreal summer and autumn). By means of this delayed ocean transport mechanism, ENSO provides an additional heat supply favourable for the formation of strong hurricanes. This thermal control on intense TC variability has significant implications for seasonal predictions and long-term projections of TC activity over the eastern North Pacific.

Tropical cyclones in the central and eastern North Pacific can have drastic economic consequences on the southwestern United States, Mexico, the Hawaiian Islands, and military and commercial maritime routes between these areas⁸. For example, Hurricane Raymond, which was quickly followed by Manuel in September 2013, triggered deadly mudslides and left about 40,000 tourists stranded in Acapulco, a popular tourist destination. In September 1992, Hurricane Iniki caused damage costing US\$2.5 billion and was the costliest hurricane to strike the state of Hawaii⁹. Although major TCs have spared Hawaii since then, a warming

environment is projected to promote an increase in these extreme events^{7,10,11}.

ENSO modulates TC activity in this region^{9,12–16}. Changes in anomalies in sea surface temperature (SST) induced by ENSO can alter atmospheric conditions, such as atmospheric stability, vertical wind shear and relative humidity^{4,5,12–14}. A recent study suggests that the strength of the cross-equatorial pressure gradient in the eastern Pacific determines the latitudinal position and intensity of the mean Intertropical Convergence Zone and its subsequent sub-weekly variability, which is responsible for the amplification of deep convection and the formation of tropical depressions¹⁶. Although previous research^{13–15} has suggested that ENSO SST and thermocline depth affect TC interannual variability over this region statistically, the physical linkage between ENSO, ocean subsurface heat and TC activity has not been established.

So far, the potential thermal impact of ENSO heat discharge on TC activity in this region has been overlooked despite the well established findings from two lines of research: first, the equatorial upper ocean heat is discharged into off-equatorial regions in the main thermocline underneath the surface as an aftermath of El Niño events^{17–19}; and second, strong TCs are capable of pumping up water from more than 100 m below the surface^{20,21}. Here we combine oceanic subsurface conditions from ORA-S3 (the reanalysis from the European Centre for Medium Range Weather Forecast)²² and TC data derived from the best track archives of the National Oceanic and Atmospheric Administration's (NOAA's) Tropical Prediction Center (TPC)²³, to demonstrate that the redistribution of subsurface heat over the top 105 m (T105) during El Niño drives the interannual variability of intense TCs over the eastern north tropical Pacific. Simply put, the well-known El Niño recharge–discharge paradigm^{17,19,24} is at the heart of ENSO's direct thermal control on intense hurricanes activity in this region.

Figure 1 shows the first two modes of variability of T105, which are defined as ocean temperature anomalies averaged between 5 and 105 m

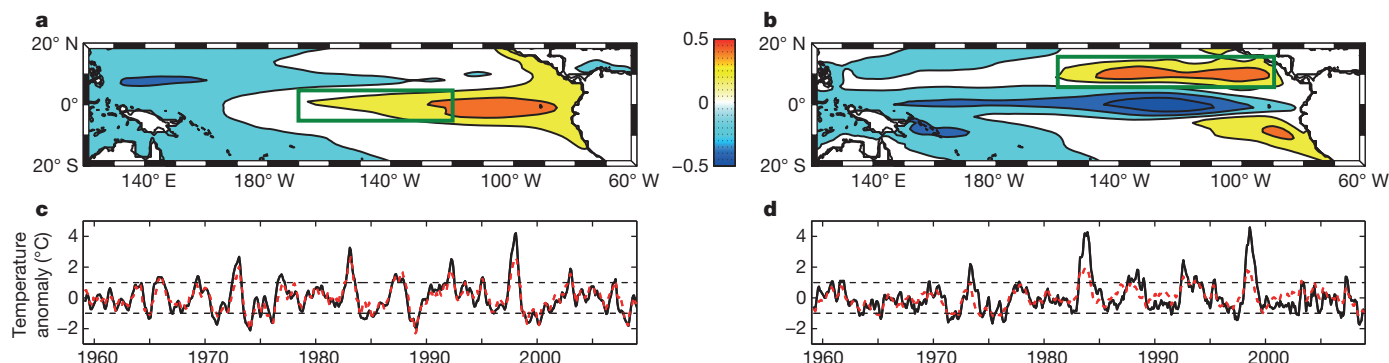


Figure 1 | EOF decomposition of T105 (temperatures averaged between 5 and 105 m) anomalies. **a, b**, The spatial patterns of the first (**a**) and second (**b**) EOF modes with explained variance at 46.8% and 11.7% respectively. **c**, The first PC (PC1, black) and the Niño3.4 time series (red) (SST anomalies averaged in the green box in **a**). **d**, The second PC (PC2, black) and time series

(red) of T105 anomalies averaged in region TC (160–90° W, 7–17° N) (green box in **b**). Horizontal dashed lines in **c** and **d** indicate ± 1 s.d. of the respective PCs. Correlation of PC1 and Niño3.4 = 0.92 ($P < 0.001$); correlation of PC2 and T105 = 0.75 ($P < 0.001$). PC time series are normalized, but units for Niño3.4 and T105 are temperature anomalies in degrees Celsius.

¹Department of Atmospheric Sciences, SOEST, University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA. ²Department of Atmospheric Sciences, National Taiwan University, Taipei 10617, Taiwan.

³Laboratory for Climate Studies, Beijing Climate Center, Chinese Meteorological Agency, Beijing 100081, China.

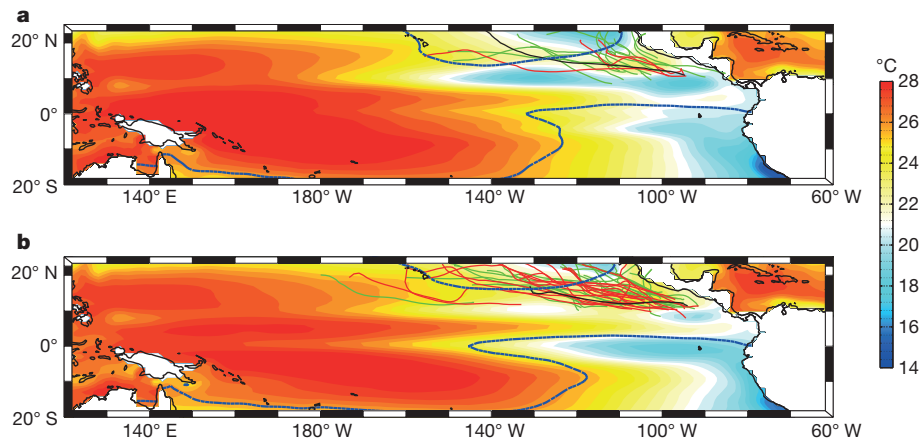


Figure 2 | Trajectories of intense TCs for periods of high and low subsurface temperature in the eastern North Pacific. The shading denotes the summer (June–November) subsurface temperature (T105) averaged over periods when $PC2 < -s.d.(PC2)$ (a) and when $PC2 > s.d.(PC2)$ (b). The thick blue line denotes the 26°C isotherm of SST averaged over the periods of low (a) and high (b) PC2 activity. In each panel the coloured lines represent the trajectories of strong TCs that occurred during the corresponding periods: green, category 3; red, category 4; black, category 5.

(units are in degrees Celsius), derived from a classic statistical decomposition into empirical orthogonal functions (EOFs). The spatial pattern of the first mode (Fig. 1a) shows the classic zonal seesaw along the Equator, which is known to be in quasi-balance with the equatorial zonal wind stress anomalies induced by ENSO SST anomalies¹⁹. Therefore, as expected, the associated time series or principal component (PC) plotted in black in Fig. 1c is directly related to the classic ENSO Niño3.4 index (that is, averaged SST anomalies in the region 170°W – 120°W , 5°S – 5°N denoted by the green box in Fig. 1a and the red dashed line in Fig. 1c) with a strong correlation of 0.92. The second EOF mode highlights the subsurface off-equatorial heat redistribution associated with ENSO (Fig. 1b) and is highly correlated with the warm water volume measuring the ENSO recharge–discharge mechanism^{19,24}. It should be noted that the prominent centre in the eastern north tropical Pacific under the region 160°W – 90°W , 7°N – 17°N (green box in Fig. 1b, referred to hereafter as region TC) has been largely missed in earlier studies of ENSO in which the ocean heat content is typically calculated as the ocean temperature averaged within the top 300 m (ref. 24) (Supplementary Fig. 1). Our studies show that T105 is much more relevant in capturing the ENSO discharge occurring in the eastern tropical Pacific, which is a highly stratified oceanic region characterized by a shallow thermocline^{25,26}. Moreover, this specific fraction of the upper subsurface temperature serves as the best potential heat reservoir fuelling the overlying TC activity^{15,20,21,26}. In addition, the second PC (black line in Fig. 1d) lags the first by about

7–8 months (lag correlation of 0.73; Supplementary Fig. 2) and is strongly correlated (0.75) with T105 anomalies averaged in region TC (red dashed line in Fig. 1d). El Niño events usually peak around Christmas time²; warm T105 anomalies discharged from the Equator as the aftermath of El Niño events will therefore peak during the following boreal summer and autumn, just in time for the active hurricane season in the Northern Hemisphere.

To highlight the link between ENSO-induced variations of T105 and TC activity in the eastern North Pacific, we contrast the records of TCs that have developed in this basin during periods of high and low T105 inferred from the second EOF mode. We define the high and low T105 periods as being when the PC2 time series is respectively above and below one standard deviation. Figure 2 shows the mean T105 and trajectories for TCs above category 3 (according to the Saffir–Simpson hurricane scale; see <http://www.aoml.noaa.gov/general/lib/laescae.html> for a description) for both composite periods. These plots indicate a clear tendency towards increased occurrences of intense TCs during anomalously warm T105 periods in the central to eastern North Pacific. This clearly suggests that T105 exerts a strong control on intense TC activity in this region. More details on how variations in T105 can influence the seasonal activeness of TC in this region can be found in Supplementary Figs 3–9 and Supplementary Tables 1 and 2.

This control on TCs by ENSO can be further delineated by examining relations between TC activity measured by the accumulated cyclone

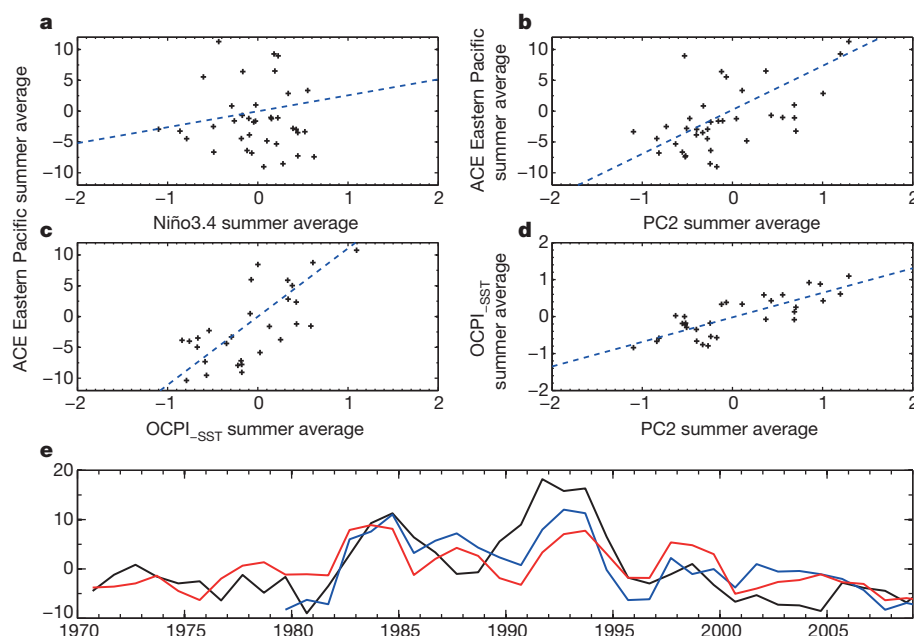


Figure 3 | Relationships between subsurface thermal conditions and TC activity in the north Eastern Pacific. a–d, Linear regressions (dashed blue lines) of boreal summer mean (June to November) ACE anomalies onto Niño3.4 (a; correlation $r = 0.15$), PC2 (b; $r = 0.62$) and OCPI_{SST} (c; $r = 0.76$), and of the summer mean of OCPI_{SST} onto PC2 (d; $r = 0.81$). Only the correlations in b, c and d are significant at a 99% confidence level. e, Time series from the bilinear regressions of summer-mean ACE (black line) onto Niño3.4 and OCPI_{SST} (blue line; $r = 0.77$) and onto Niño3.4 and PC2 (red line; $r = 0.64$).

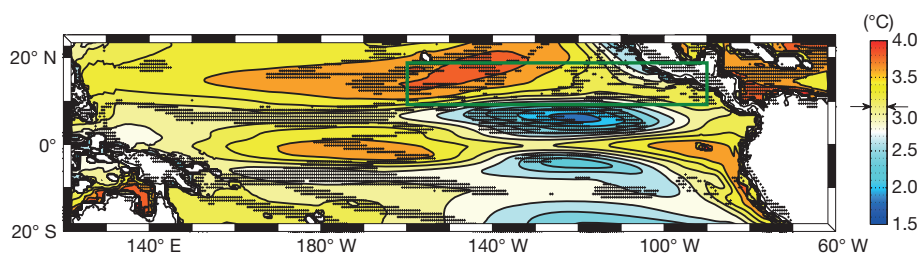


Figure 4 | Changes in subsurface T105 between the twentieth-century historical conditions and the rcp8.5 global warming scenario. Multi-model ensemble mean of mean T105 difference between the rcp8.5 scenario and historical simulations for models indicated in Supplementary Table 3. The two arrows on each side of the colour bar indicate the mean T105 difference for the basin. The green box represents region TC (160–90°W, 7–17°N). The cross-hatching denotes a statistical confidence level of 99% based on Student's *t*-test.

energy index (ACE) in the central to eastern North Pacific²⁷, a SST-based potential intensity index (SSTPI) and upper-ocean heat-content-based ocean-coupled potential intensity index (OCPI) for TCs²¹, and PC1/PC2 time series. ACE measures the total kinetic energy of an observed individual hurricane by convolving the frequency, duration and intensity of the TC from best track archives^{23,27} and thus emphasizes the strength of TC activity. SSTPI and OCPI are two theoretical indices that estimate hurricane potential intensities under known local background atmospheric and ocean thermal conditions²¹. We design a simple bilinear regression model of ACE anomalies onto PC1 and PC2, $ACE = aPC_1 + bPC_2$. We also linearly regress ACE onto each of the individual PCs to contrast the importance of these two factors. For seasonal anomalies averaged over the entire active season for hurricanes (that is, June to November), this simple regression model yields a correlation of 0.66. When correlated with PC1 and PC2 individually, the correlations are about 0.20 and 0.62, respectively (Supplementary Fig. 10). This indicates the unambiguous dominance of the delayed ENSO thermal effect on TC activity through El Niño subsurface heat discharge.

Figure 3 further examines the relationship between ACE, Niño3.4, PCs and SSTPI/OCPI and attempts to reconcile these different indices. The latter two theoretical indices are calculated directly in region TC. It should be noted that SSTPI and OCPI are not entirely independent because SST contributes to OCPI throughout the mixed layer. We express $ACE = a_1Niño3.4 + b_1OCPI_{-SST}$, where $OCPI_{-SST} = OCPI - cSSTPI$ accounts only for the contribution of subsurface thermal information. We also consider another simple bilinear regression model for completeness, $ACE = a_2Niño3.4 + b_2PC_2$. In this way, the first and second terms of both bilinear regressions reflect the contribution to ACE from SST anomalies and subsurface temperature anomalies, respectively. The first term measures the ability of ENSO SST anomalies to alter key atmospheric factors for TC genesis^{4,5,12–16}. Again we find that the second term dominates the connection with ACE; the correlations between ACE and $OCPI_{-SST}$ alone and between ACE and PC2 alone reach 0.76 and 0.62, respectively (Fig. 3b, c). Moreover, $OCPI_{-SST}$ is highly related to PC2 at a correlation of about 0.81 (Fig. 3d), whereas OCPI itself is less correlated with PC2 as a result of noise contaminations from local SST anomalies. Because ACE is skewed towards intense TCs, we conducted additional investigations (including the analyses of atmospheric model outputs; see Supplementary Information), to break down the role of the direct (PC1) and 'stealth' (PC2) ENSO control into, respectively, moderate and intense TC activity. PC1 and moderate storm activity are somewhat related, which is in agreement with theories that attribute the genesis of tropical depressions to ENSO-induced atmospheric disturbances. However, these extra diagnostics confirm a much more robust relationship between the occurrences of intense hurricanes and the El Niño subsurface heat variability by PC2. This further supports the notion that the ocean thermal effect of ENSO through its meridional subsurface heat redistribution is the dominant factor that controls yearly variations of extreme TC activity in the central to eastern North Pacific. Strong El Niño events provide subsurface energy during the subsequent boreal summers for TCs to develop into major hurricanes.

When measured by the Niño3.4 index, ENSO has been shown to be highly predictable, usually about two to three seasons ahead²⁸. The T105 discharge reaches a maximum about three seasons after the peak of El

Niño. Observed ENSO signals in the winter are thus good indicators of TC activity during the subsequent summer in the central to eastern North Pacific, with the potential to capture about 40–70% of the yearly ACE variability. Because of the environmental and societal impacts of intense hurricanes, and even though the individual TC tracking still remains a considerable challenge, this high predictability of extreme hurricane activity may be valuable for surrounding regions.

This insight into the strong influence of ENSO over TC activity in the eastern Pacific has implications on how the hurricane intensity in this region may respond to global warming. The global climate is projected to warm up significantly as a result of increasing anthropogenic forcing²⁹. The most recent climate simulations from the Coupled Model Intercomparison Project version 5 predict a strongly enhanced subsurface warming pattern in the eastern North Pacific under increased concentrations of greenhouse gases (Fig. 4 for the rcp8.5 scenario). This warming, while subject to the uncertainties of model simulations, has a local maximum located just underneath the prevailing historical TC route (region TC shown in Fig. 4). Although subject to further uncertainties, such as in upper tropospheric temperatures, this suggests an increased probability, estimated to be between 25% and 45% compared with present climatic conditions, for more intense hurricanes in this region in the future (Supplementary Fig. 20).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 April; accepted 9 October 2014.

- Glantz, M. H. *Current of Change: Impacts of El Niño and La Niña on Climate and Society* (Cambridge Univ. Press, 2001).
- McPhaden, M. J., Zebiak, S. E. & Glantz, M. H. ENSO as an integrating concept in earth science. *Science* **314**, 1740–1745 (2006).
- Gray, W. M. in *Meteorology over the Tropical Oceans* (ed. Shaw, D. B.) 155–218 (Royal Meteorological Society, 1979).
- DeMaria, M. The effect of vertical shear on tropical cyclone intensity change. *J. Atmos. Sci.* **53**, 2076–2087 (1996).
- Dong, K. & Holland, G. J. A global view of the relationships between ENSO and tropical cyclone frequencies. *Acta Met. Sin.* **8**, 19–29 (1994).
- Wang, C. & Lee, S. Co-variability of tropical cyclones in the north Atlantic and the eastern North Pacific. *Geophys. Res. Lett.* **36**, L24702 (2009).
- Peduzzi, P. B. et al. Global trends in tropical cyclone risk. *Nature Climate Change* **2**, 289–294 (2012).
- Court, A. *Tropical Cyclone Effects on California* (National Oceanic and Atmospheric Administration Tech. Memo. NWS WR-159, 1980).
- Chu, P. S. & Wang, J. Tropical cyclone occurrences in the vicinity of Hawaii: are the differences between El Niño and non-El Niño years significant? *J. Clim.* **10**, 2683–2689 (1997).
- Murakami, H., Wang, B., Li, T. & Kitoh, A. Projected increase in tropical cyclones near Hawaii. *Nature Climate Change* **3**, 749–754 (2013).
- Webster, P. J., Holland, G. J., Curry, J. A. & Chang, H. R. Changes in tropical cyclone number, duration and intensity in a warming environment. *Science* **309**, 1844–1846 (2005).
- Camargo, S. J., Sobel, A. H., Barnston, A. G. & Ghil, M. Clustering of eastern north Pacific tropical cyclone tracks: ENSO and MJO effects. *Geochem. Geophys. Geosyst.* **9**, Q06V05 (2008).
- Collins, J. M. & Mason, I. M. Local environmental conditions related to seasonal tropical cyclone activity in the northeast Pacific basin. *Geophys. Res. Lett.* **27**, 3881–3884 (2000).
- Collins, J. M. The relationship of ENSO and relative humidity to interannual variations of hurricane frequency in the north-east Pacific Ocean. *Pap. Appl. Geogr. Conf.* **30**, 324–333 (2007).
- Balaguru, K., Leung, R. L. & Yoon, J.-H. Oceanic control of northeast Pacific hurricane activity at interannual timescales. *Environ. Res. Lett.* **8**, 044009 (2013).

16. Toma, V. & Webster, P. J. Oscillations of the intertropical convergence zone and the genesis of easterly waves. Part I: diagnostics and theory. *Clim. Dyn.* **34**, 587–604 (2010).
17. Wyrtki, K. Water displacements in the Pacific and the genesis of El Niño cycles. *J. Geophys. Res. Oceans* **90**, 7129–7132 (1985).
18. Jin, F. F. Tropical ocean–atmosphere interaction, the Pacific cold tongue, and the El Niño Southern Oscillation. *Science* **274**, 76–78 (1996).
19. Jin, F. F. An equatorial ocean recharge paradigm for ENSO. Part I: conceptual model. *J. Atmos. Sci.* **54**, 811–829 (1997).
20. Price, J. F. Metrics of hurricane–ocean interaction: vertically-integrated or vertically-averaged ocean temperature? *Ocean Sci.* **5**, 351–368 (2009).
21. Lin, I. I. *et al.* An ocean coupling potential intensity index for tropical cyclones. *Geophys. Res. Lett.* **40**, 1878–1882 (2013).
22. Balmaseda, M. A., Vidard, A. & Anderson, D. L. T. The ECMWF Ocean Analysis System: ORA-S3. *Mon. Weath. Rev.* **136**, 3018–3034 (2008).
23. NOAA's Tropical Prediction Center. <http://www.nhc.noaa.gov/?epac>.
24. Meinen, C. S. & McPhaden, M. J. Observations of warm water volume changes in the equatorial Pacific and their relationship to El Niño and La Niña. *J. Clim.* **13**, 3551–3559 (2000).
25. Clement, A. C., Seager, R., Cane, M. A. & Zebiak, S. E. An ocean dynamical thermostat. *J. Clim.* **9**, 2190–2196 (1996).
26. Shay, L. K. & Brewster, J. K. Oceanic heat content variability in the Eastern Pacific Ocean for hurricane intensity forecasting. *Mon. Weath. Rev.* **138**, 2110–2131 (2010).
27. Maue, R. N. Recent historically low global tropical cyclone activity. *Geophys. Res. Lett.* **38**, L14803, <http://dx.doi.org/10.1029/2011GL047711> (2011).
28. Latif, M. D. *et al.* A review of the predictability and prediction of ENSO. *J. Geophys. Res. Oceans* **103**, 14375–14393 (2012).
29. Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) (Cambridge Univ. Press, 2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Kim for providing the CMIP5 model's outputs; M. Zhao for making his Atmospheric General Circulation Model outputs available; K. Emanuel for providing the potential intensity calculation program via his website; and C. Huang for help in data processing. This study was supported by US National Science Foundation grants ATM1034798 and ATM1406601, US Department of Energy grant DESC0051110, US NOAA grant NA100AR4310200, the China Meteorological Special Project (GYHY201206033) and the 973 Program of China (2010CB950404 and 2013CB430203). I.I.L.'s work is supported by Taiwan's Ministry of Science and Technology under grants NSC 101-2111-M-002-002-MY2, NSC 101-2628-M-002-001-MY4 and 102R7803.

Author Contributions F.F.J. conceived the idea. J.B. conducted most of the analysis. I.I.L. contributed to SSTPI and OCPI calculations and analysis. F.F.J. and J.B. contributed to writing the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.-F.J. (jff@hawaii.edu) or J.B. (bouch@hawaii.edu).

METHODS

The oceanic features of ENSO are represented using an EOF analysis of the ocean thermal information in the tropical Pacific³⁰. The first and second modes represent the zonal and meridional swings, respectively, of the warm water volume associated with the recharge–discharge theory^{17,19,24}. Then we use the second mode PC (in time series) to separate and compose periods of strong off-equatorial warmth recharge from discharge in the northeastern Pacific. Finally, we compile TC trajectories from the best track archives of NOAA's Tropical Prediction Center during these distinct composite periods to assess the role of meridional heat redistribution on TC activity in the eastern Pacific²³. The region of interest in this article (that is, region TC) is defined as the area of maximum subsurface anomaly warming in the eastern North Pacific as inferred from the second mode of the T105 EOF decomposition. This maximum warming collocates with the main region of TC formation in the Eastern Pacific.

Below we present basic documentation about the data and indices used in this study to characterize TC activity.

ORA-S3. Details can be found in http://adprc.soest.hawaii.edu/dods/public_data/Reanalysis_Data/ORA-S3.

Tropical cyclone tracks. Details can be found in <http://www.nhc.noaa.gov/?epac>. **SSTPI and OCPI calculation.** Potential intensity (PI) is an important fundamental concept and estimation of the upper bound to the intensity of a TC^{31–33}. In this study, the PI calculations are based on the program provided by K. Emanuel (Massachusetts Institute of Technology) from his website, using input from the European Centre for Medium-Range Weather Forecasts (ECMWF):

$$V^2 = \frac{SST - T_0}{T_0} \frac{C_k}{C_D} (k^* - k) \quad (1)$$

A TC's maximum intensity (maximum surface wind speed, V) is estimated as a function of SST (hence the name SSTPI in this study), T_0 , which is the TC outflow temperature determined by the atmospheric vertical profile; the drag coefficient C_D , the enthalpy exchange coefficient C_k , the saturation enthalpy k^* of the sea surface, and the surface enthalpy k in the TC environment. As in equation (1), the original PI (that is, SSTPI) uses only SST to represent the ocean's contribution to PI. However, the thermal condition of the upper subsurface ocean column (typically from the surface down to ~100 m) is also crucial to a TC's intensification^{34–38}. A recent proposal used the upper ocean temperature to replace SST to form OCPI²¹, which is a revised index. OCPI replaces SST with the depth-averaged ocean temperature \bar{T} averaged from the surface down to the expected TC–ocean coupling depth, typically about 60–120 m (refs 20, 21), to calculate PI:

$$V_{OCPI}^2 = \frac{\bar{T} - T_0}{T_0} \frac{C_k}{C_D} (k^* - k)$$

By examining 14 years (1998–2011) of TC cases over the western North Pacific, it has been found that OCPI is more closely related to the observed TC maximum intensity than to SSTPI²¹. The atmospheric input was from the monthly atmospheric temperature and humidity profile of the ECMWF Interim Reanalysis database at each 1° grid. The SST and \bar{T} (that is, T105, upper 105 m average) are also on a 1° grid and based on the monthly ocean reanalysis data of the ECMWF (ORAS3) from 1979 to 2009.

Accumulated cyclone information calculation. Accumulated cyclone information (ACE)²⁷ is a measure used by NOAA to express the activity of individual TCs and entire TC seasons. It uses an approximation of the total kinetic energy of a TC over its lifetime and is calculated every 6 h. The ACE of a season is the sum of the ACEs for each storm and takes into account the number, intensity and duration of all the TCs. The sustained wind speed of every active TC of above tropical storm intensity (wind speed 35 knots (65 km h^{−1}); that is, Tropical Storm category or higher), is also calculated at 6-h intervals. If any storms during a season happen to cross years, the storm's ACE counts for the previous year. The unit of ACE is 10⁴ kt² and this is the unit used for the index. Thus, $ACE = 10^{-4} \sum v_{max}^2$, where v_{max} is the estimated sustained wind speed in knots.

The ACE index data come from the NOAA's Climate Prediction Center, available online at http://www.cpc.noaa.gov/products/outlooks/background_information.shtml.

Analyses on ENSO control on TC activity in the eastern North Pacific basin. In contrast with the first two EOFs of T105 shown in Fig. 1, Supplementary Fig. 1 shows the first two EOF modes of the subsurface temperature monthly anomalies averaged between 5 and 305 m (T305 hereafter), a commonly used averaged depth in existing ENSO climate studies. It illustrates that such a deep averaging significantly diminishes the prominent centre of shallower heat content variability in the eastern north tropical Pacific (T105 averaged in region TC (160–90° W, 7–17° N); Fig. 1b).

Supplementary Fig. 2 shows the autocorrelation and lag correlation of the first two PCs from the EOF analysis performed on T105 (averaged between 5 and 105 m).

In addition to the result shown in Fig. 2, Supplementary Fig. 3 contrasts the strong TC (categories 4 and 5 only) development between periods of high and low T105 (as in Fig. 2, but only hurricanes of categories 4 and 5 are shown). Note that these composite periods cover 81 and 66 months, respectively. The difference in track density is even more striking than on Fig. 2. Supplementary Fig. 4 is the same as Supplementary Fig. 3 but contrasts periods of low and high T105 inferred from PC1 threshold. It highlights that the ENSO indirect control (through atmospheric disturbances triggered by ENSO-induced SST anomalies; see the text) is much less significant in driving strong hurricane development than the direct but delayed control of meridional subsurface heat redistribution (T105/PC2). The top panels of Supplementary Figs 5 and 6 are the same as Supplementary Figs 3 and 4, but the trajectories of all categories of storms are plotted (ranging from tropical depression to category 5 hurricane). The middle panels show the mean SST (left) and T105 (right) difference between periods of high and low values of PC2 (Supplementary Fig. 5) and PC1 (Supplementary Fig. 6). The bottom panels of these figures show the PC time series and standard deviations used as thresholds to identify the high and low value periods. Subsurface ocean temperature is significantly warmer during high PC2 periods underneath the TC formation region, whereas SST does not show much change (Supplementary Fig. 5). Conversely, the subsurface temperature is slightly smaller during the periods of high PC1 in region TC, whereas SST does not show much change in this area between periods of high and low PC1. The TC activity in the eastern North Pacific is therefore predominantly associated with the oceanic subsurface variability related to the second mode of variability of heat content (PC2 of T105). This confirms that the T105 variation associated with meridional subsurface heat exchange is the main driver of strong hurricane activity in this region. Supplementary Fig. 7 depicts a different approach, in which we composed periods of high and low TC activity (evaluated using the ACE index) in the eastern North Pacific (in a similar manner to that used previously) and evaluated the corresponding mean T105 and SST patterns associated with these distinct periods. The differences shown in Supplementary Fig. 7 unambiguously associate a strong subsurface warming in region TC with anomalously active hurricane periods; again, we see only small SST differences in region TC.

Supplementary Table 1 provides the total number of each category of TC that developed during the composite periods identified using both PC1 and PC2 thresholds (from T105 EOF decomposition). Because it is nearly identical to Niño3.4, PC1 identifies itself with the indirect ENSO effect on TC activity through atmospheric controls. It actually measures the ability of ENSO to 'alter' key atmospheric factors involved in TC generation and growth, such as atmospheric stability, vertical wind shear and relative humidity. Having not received much attention yet, PC2 represents the direct thermal impact of ENSO on TC activity in this region through the delayed redistribution of subsurface meridional heat. We note a 74% versus 42% increase in the total number of TCs between high and low PC2 versus PC1 periods. However, this contrast of percentage change becomes 350% (100 × (18 − 4)/4) versus 43% (100 × (10 − 7)/7) for strong TC (categories 4 and 5; Supplementary Figs 3 and 4) between high and low PC2 versus PC1 periods. This direct thermal control of ENSO on TC activity through delayed subsurface heat discharge is thus clearly effective, particularly for the development of intense hurricanes in the central to eastern North Pacific. Supplementary Fig. 8 is a bar diagram representation of Supplementary Table 1. Supplementary Table 2 is similar but uses the Niño3.4 index and T105 (that is, subsurface temperature averaged in region TC) instead of PC1 and PC2. We note 45% and 14% increases in the total number of TCs between the high and low T105 period for the T105 index and Niño3.4 criteria, respectively. There are 350% and 25% increases in strong TCs (categories 4 and 5) between the high and low T105 period for the T105 index and Niño3.4 criteria, respectively. Finally, there are 144% and 71% increases in TCs (categories 3, 4 and 5) between the high and low T105 period for the T105 index and Niño3.4 criteria, respectively. Supplementary Fig. 9 is a bar diagram representation of Supplementary Table 2.

In addition to the result shown in Fig. 3, we also design a simple bilinear regression model of ACE monthly anomalies onto PC1 and PC2, $ACE = aPC_1 + bPC_2$ (Supplementary Fig. 10c). We also linearly regress ACE onto each of the individual PCs to contrast the importance of these two factors (Supplementary Fig. 10a, b). For seasonal anomalies averaged over the entire hurricane active season (that is, June to November), this simple regression model yields a correlation of 0.66 (Supplementary Fig. 10c), whereas when correlated with PC1 and PC2 individually the correlations are about 0.20 (Supplementary Fig. 10b) and 0.62 (Supplementary Fig. 10a) for PC1 and PC2, respectively.

Distinction between ENSO's controls on intense and weak TCs. Because the ACE index is skewed towards more intense TCs, we provide additional analysis to break down the role of both the direct (PC1/Niño3.4) and 'stealth' (PC2) ENSO controls onto weak and intense TC activity.

To distinguish clearly between frequency and intensity in our data, we first examined the relationships between a classic count of all hurricanes that developed in the eastern Pacific, using data from NOAA's TPC best track data site and either PC1 (or Niño3.4) or PC2 time series inferred from the EOF analysis of T105 from observations (ORA-S3 reanalysis). We then performed a similar diagnostic using only intense TCs (category 3 and above) and relatively weak TCs or moderate storms (tropical depressions to category 2 hurricanes). Again, we assume that the PC1/Niño3.4 index represents the indirect ENSO effect on TCs and is associated with large-scale SST variations that can trigger atmospheric disturbances (for example, variations in the position and strength of the Intertropical Convergence Zone). In contrast, PC2 measures our stealth ENSO effect associated with delayed meridional variations of subsurface oceanic heat.

Supplementary Fig. 11 shows the annual count of all storm categories according to the Saffir–Simpson scale over the whole ORA-S3 period (1959–2009). To focus on the interannual variability, we detrended all time series.

The regressions between the annual TC count and the different indices (PC1/Niño3.4 representing the direct ENSO control, and PC2 representing the stealth ENSO control) for the entire ORA-S3 reanalysis period (1959–2009) are shown in Supplementary Fig. 12. The relationship between the total annual count and both Niño3.4 (Supplementary Fig. 12a) and PC2 (Supplementary Fig. 12b) does not seem to be significant. However, when only intense TCs are considered (category 3 and above), the correlation between the hurricane count and PC2 (Supplementary Fig. 12d) increases significantly, whereas it stays within the same value with Niño3.4 (Supplementary Fig. 12c). In contrast, there is no clear link between ENSO and moderate tropical storms.

To give additional support to the link between PC2 and intense TC variability, we conducted similar analyses using outputs from a recent Atmospheric General Circulation Model (AGCM) modelling study³⁹. Simulations with finer atmospheric resolutions (20–100 km) are useful for assessing many aspects of storms genesis, such as distribution and even interannual variability of storm frequency, although the simulations of storm intensity still remain highly inadequate, particularly in the eastern Pacific^{40–43}. To capture and assess the direct relationship between ENSO (PC1) and TC frequency, four integrations of a relatively high-resolution AGCM forced by SSTs specified from the UK Met Office Hadley Centre Sea Ice and SST version 1.1 model (HadISST 1.1 (ref. 44)) were performed in the study³⁹. They tracked tropical storms with a wind speed greater than 17 m s^{-1} for three consecutive days but did not differentiate between hurricane categories. In addition, it is worth noting that the probability distribution function (PDF) of the eastern Pacific simulated storm intensity (as diagnosed from the maximum wind speed attained; Fig. 6 of ref. 39) shows serious biases compared with the observed PDF, especially for wind speeds greater than 50 m s^{-1} . In other words, their AGCM does not simulate storms stronger than category 2 hurricanes in the eastern Pacific.

Supplementary Fig. 13 shows the total annual count of TCs generated by the four experiments and their ensemble mean. The timing of the hurricane season is well reproduced by the model (active months are June to November), whereas the interannual variability in TC frequency in the Eastern Pacific (Fig. 9 of ref. 39) is of only limited reproducibility. The regressions between the total count of TCs and Niño3.4 extracted from the SST forcing field or PC2 from ORA-S3 are shown in Supplementary Fig. 14. The regression between TC counts and Niño3.4 does show a significant correlation in the simulations. As expected, there is little correlation with the PC2 time series because there is no a priori information from the subsurface in the AGCM experiments. Supplementary Fig. 15 shows the regression for the same period as the AGCM experiments (1981–2009) from observations. This period shows the same insignificant correlation between Niño3.4 and total TC count, similar to the result for the entire period (Supplementary Fig. 12). The correlation becomes even smaller for intense TCs and Niño3.4, but there is a weak correlation for moderate tropical storms, at a level comparable to the AGCM results. However, for this most recent period, the correlation between PC2 and intense hurricane counts is statistically significant (~ 0.7). Thus, a weak indirect influence of ENSO on TC activity through atmospheric responses is largely related to relatively weak TC counts, whereas the stealth and delayed subsurface ocean thermal effect we uncovered in this study is very robustly linked to intense TC activity.

A long-term projection of extreme TC activity in the eastern North Pacific basin. Supplementary Table 3 and Supplementary Figs 16–20 complete Fig. 4 and provide additional insights on how TC activity in this region may be affected by global warming. The former presents the multi-model ensemble mean of the SST difference between rcp8.5 and historical scenarios⁴⁵ for models shown in Supplementary Table 3. Supplementary Fig. 17 is the same as Fig. 4 but uses a different statistical test to emphasize the statistically confident warming patterns (see the captions for a description of the statistical tests). Supplementary Fig. 18 shows the multi-model ensemble mean of the mean SST difference between rcp4.5 and the historical runs. Supplementary Fig. 19 is the same as Fig. 4 but uses the rcp4.5 scenario instead of rcp8.5. Supplementary Fig. 20a (top panel) is the same as Fig. 4 except that the

basin mean, which is speculated to approximate the uniform atmospheric warming to TC outflow temperature. To obtain a rough estimate of the possible increase in intense TCs between the warm and historical climatic conditions over the tropical Pacific, we contrast the number of TCs per decade during low and high T105 and the T105 anomalies averaged in the region shown in Supplementary Fig. 20a. If we assume the regional enhanced warming relative to the basin mean as an anomaly, we can extrapolate the relationships between T105 anomalies and the number of TCs per decade to warming climate scenarios. The multi-model ensemble projects an enhanced warming of 0.43°C in the eastern Pacific TC formation region, which is delineated by the thick dashed green line in Supplementary Fig. 20a. This corresponds to ~ 46 TCs per decade and represents a 43% increase in intense hurricanes between current climatic conditions and the most severe global warming scenario (rcp8.5) in the central to eastern North Pacific. This increase between current climatic conditions and the rcp4.5 scenario is about 24%. The study³⁹ suggested an increase of around 40% for the total number of TCs (only moderate ones, because their AGCM failed to simulate intense hurricanes) between current climatic conditions and the most severe global warming scenario (A1B from the CMIP3 database) in the eastern Pacific. Note that the increase in the normalized (per decade) number of category 3 and above TCs between low and high T105 historical periods (inferred from PC2 time series threshold) is 87% for this region.

Statistical significance of correlations. PC2 and all ENSO-related indices are strongly autocorrelated, which significantly reduces the denominator of degrees of freedom (ddf) of the time series. To provide more statistical confidence on the correlation coefficients calculated throughout the paper and the Supplementary Information, we reassess the *t*-test PDF values with a revised ddf. To do so, we take into account the decorrelation timescales of Niño3.4 and the PC2 time series to assess an 'effective' ddf. First, we compute the autocorrelation of both Niño3.4 and PC2 monthly and yearly time series that span the different periods of time diagnosed in this study (compare Supplementary Figs 21 and 22, in the same way as Supplementary Fig. 2). For monthly values (time series of $N = 612$ samples for the total period 1959–2009), the time series are not significantly autocorrelated after 8–11 months depending on the index and time period. This reduces the ddf to $N = 41–77$ (see Supplementary Tables 1 and 2). Similarly, for the boreal summer average the ddf is reduced to $N = 21–34$. Note that this is a really drastic criterion because the annual time series are hardly autocorrelated. On the basis of these new ddfs, we can now calculate the correlation values for which the *t*-distribution probability is below 0.0001 (99% statistical confidence). These values are indicated in Supplementary Tables 4 and 5 for Niño3.4 and PC2, respectively, for the different periods used in this study; they give the lowest values for a correlation coefficient that is significant at a 99% confidence level. Supplementary Table 6 provides the regular *P* values (calculated with the initial ddf) of correlations calculated throughout the paper and the Supplementary Information.

Miscellaneous. Supplementary Fig. 23 provides insights into the temperature vertical profile in region TC during high and low T105 periods (inferred from the PC2 time series threshold). Strong subsurface variability is confined beneath the surface (between 30 and 100 m). The low T105 periods show an enhanced thermocline ridge around 11°N . This brings up anomalously cold water in this region that is likely to prevent TCs from developing into major hurricanes during these low T105 periods.

30. Lorenz, E. N. *Empirical Orthogonal Functions and Statistical Weather Prediction* (Massachusetts Institute of Technology Department of Meteorology, Statistical Forecasting Project Report 1, 1956).
31. Emanuel, K. A. The maximum intensity of hurricanes. *J. Atmos. Sci.* **45**, 1143–1155 (1988).
32. Emanuel, K. A. Sensitivity of tropical cyclones to surface exchange coefficients and a revised steady-state model incorporating eye dynamics. *J. Atmos. Sci.* **52**, 3969–3976 (1995).
33. Bister, M. & Emanuel, K. A. Dissipative heating and hurricane intensity. *Meteorol. Atmos. Phys.* **52**, 233–240 (1998).
34. Leipper, D. F. & Volgenau, D. Hurricane heat potential of the Gulf of Mexico. *J. Phys. Oceanogr.* **2**, 218–224 (1972).
35. Shay, L. K., Goni, G. J. & Black, P. G. Effects of a warm oceanic feature on Hurricane Opal. *Mon. Weath. Rev.* **128**, 1366–1383 (2000).
36. Lin, I. I., Wu, C. C., Pun, I. F. & Ko, D. S. Upper-ocean thermal structure and the western North Pacific category 5 typhoons. Part I: ocean features and the category 5 typhoons' intensification. *Mon. Weath. Rev.* **136**, 3288–3306 (2008).
37. Goni, G. et al. Applications of satellite-derived ocean measurements to tropical cyclone intensity forecasting. *Oceanography* **22**, 190–197 (2009).
38. Shay, L. K. & Brewster, J. K. Oceanic heat content variability in the eastern Pacific Ocean for hurricane intensity forecasting. *Mon. Weath. Rev.* **138**, 2110–2131 (2010).
39. Zhao, M., Held, I. M., Lin, S. J. & Vecchi, G. A. Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *J. Clim.* **22**, 6653–6678 (2009).
40. Bengtsson, L., Hodges, K. & Esch, M. Tropical cyclones in a T159 resolution global climate model: comparison with observations and re-analyses. *Tellus* **59A**, 396–416 (2007).
41. Bengtsson, L. et al. How may tropical cyclones change in a warmer climate? *Tellus* **59A**, 539–561 (2007).

42. LaRow, T., Lim, Y.-K., Shin, D., Chassignet, E. & Cocke, S. Atlantic basin seasonal hurricane simulations. *J. Clim.* **21**, 3191–3206 (2008).
43. Oouchi, K. *et al.* Tropical cyclone climatology in a global warming climate as simulated in a 20 km mesh global atmospheric model: frequency and wind intensity analysis. *J. Meteorol. Soc. Jpn.* **84**, 259–276 (2006).
44. Rayner, R. *et al.* Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, 4407, <http://dx.doi.org/10.1029/2002JD002670> (2003).
45. Moss, R. H. *et al.* The next generation of scenarios for climate change research and assessment. *Nature* **463**, 747–756 (2010).

Business culture and dishonesty in the banking industry

Alain Cohn^{1*}, Ernst Fehr¹ & Michel André Maréchal^{1*}

Trust in others' honesty is a key component of the long-term performance of firms, industries, and even whole countries^{1–4}. However, in recent years, numerous scandals involving fraud have undermined confidence in the financial industry^{5–7}. Contemporary commentators have attributed these scandals to the financial sector's business culture^{8–10}, but no scientific evidence supports this claim. Here we show that employees of a large, international bank behave, on average, honestly in a control condition. However, when their professional identity as bank employees is rendered salient, a significant proportion of them become dishonest. This effect is specific to bank employees because control experiments with employees from other industries and with students show that they do not become more dishonest when their professional identity or bank-related items are rendered salient. Our results thus suggest that the prevailing business culture in the banking industry weakens and undermines the honesty norm, implying that measures to re-establish an honest culture are very important.

Banks and financial markets are fundamental pillars of every advanced economy, and banking services are a key requirement for economic development¹¹. The recent wave of problematic financial practices and scandals involving fraud, however, has caused serious economic damage and may threaten the stability and reputation of banks and the financial sector as a whole. For example, the practices of the two rogue traders Jérôme Kerviel and Kweku Adoboli resulted in total damages of almost US\$10 billion⁵. Other examples, such as the manipulation of key interest rates (for example, LIBOR⁶), as well as several tax evasion affairs¹², have led to a dramatic loss of reputation and a crisis of trust⁷ in the financial sector.

Policy makers are increasingly concerned about these scandals, and often attribute problems to the prevailing business culture (that is, the norms and informal rules) in the banking industry⁸, which is thought to tolerate dishonest behaviours instead of prohibiting them. Lord Adair Turner, who was chairman of the UK Financial Services Authority from 2008 to 2013, argued that “bank executives face the challenge of setting clearly from the top a culture which tells people that there are things they shouldn't do, even if they are legal, even if they are profitable and even if it is highly likely that the supervisor will never spot them”¹⁰.

Despite its policy relevance, sound empirical evidence on the culture in the banking industry is lacking. Identifying the effects of business culture poses several challenges. Simply comparing the prevalence of dishonesty among bank employees and other professionals can be misleading. More honest people may select different professions than less honest people, which makes it difficult to disentangle cultural and selection effects. Moreover, professional groups vary along many unobservable dimensions. For example, differences in honesty could be related to differences in personal financial situations across professions. Failing to account for such dimensions could result in spurious correlations.

We take a novel approach that is inspired by the economic theory of identity¹³, which proposes that individuals have multiple social identities based on, for example, gender, ethnicity or profession. Identities are associated with specific social norms prescribing permissible behaviours.

Which identity and associated norms are behaviourally relevant depends on the relative weight an individual attributes to an identity. In a given situation, behaviour is shifted towards those norms that are associated with the more salient identity^{14–17}. Thus, if the banking culture favours dishonest behaviours, it should be possible to trigger dishonesty in bank employees by rendering their professional identity salient¹⁴.

To study whether the banking culture causes dishonesty, we recruited 128 bank employees from a large, international bank for an experiment. We randomly assigned the employees either to a treatment condition that increased the salience of their professional identity or to a control condition in which their professional identity was not made salient. Random assignment ensures that subjects in the two conditions are statistically indistinguishable in all observable and unobservable characteristics. This circumvents problems related to selection and spurious correlation.

On average, the bank employees had 11.5 years of experience in the banking industry. Roughly half of them worked in a core business unit, that is, as private bankers, asset managers, traders or investment managers. The others came from one of the support units (for example, risk or human resources management). Subjects participated in a short online survey. After answering some filler questions about subjective wellbeing, subjects in the professional identity condition were asked seven questions about their professional background (for example, “At which bank are you presently employed?” or “What is your function at this bank?”). Those in the control condition were asked seven questions that were unrelated to their profession (for example, “How many hours per week do you watch television on average?”).

After the priming questions, all subjects anonymously performed a coin tossing task that has been shown to reliably measure dishonest behaviour in an unobtrusive way^{18–20} and to predict rule violation outside the laboratory¹⁷. The rules required subjects to take any coin, toss it ten times, and report the outcomes online. For each coin toss they could win an amount equal to approximately US\$20 (as opposed to \$0) depending on whether they reported ‘heads’ or ‘tails’. Subjects knew in advance whether heads or tails would yield the monetary payoff for a specific coin toss. Moreover, subjects were informed that their earnings would only be paid out if they were higher than or equal to those of a randomly drawn subject from a pilot study. We introduced this element to mimic the competitive nature of the banking profession⁹. Given that the maximum payoff is approximately \$200, subjects faced a considerable incentive to cheat by misreporting the outcomes of their coin tosses. Because subjects were unobserved during the task and because they could hide behind chance, it is impossible to identify who cheated. However, we can detect dishonesty at the group level by comparing the reported fraction of successful outcomes with the 50% benchmark implied by honest reporting. Using similar paradigms, researchers found that in situations in which cheating cannot be detected, many individuals do not take full advantage of cheating opportunities^{18–22}. This phenomenon is typically attributed to the presence of an honesty norm^{4,23} that causes psychological costs when one cheats, such as the loss of one's positive self-image²¹.

¹Department of Economics, University of Zurich, 8006 Zurich, Switzerland.

*These authors contributed equally to this work.

We conducted a manipulation check in which subjects converted word fragments into meaningful words. For example, they could complete the word fragment “__oker” with the bank-related word “broker” or an unrelated word such as “smoker”. This allowed us to test whether the treatment increased professional identity salience. The frequency of bank-related words in the professional identity condition was increased by 40%, from 26% in the control to 36% ($P = 0.035$, rank-sum test), indicating that our manipulation was successful.

Figure 1a shows the binomial distribution of earnings in the coin tossing task that would result if everyone behaved honestly, and the empirical distribution from the control condition. Both distributions closely overlap, suggesting that the control group behaved mostly honestly. On average, they reported successful coin flips in 51.6% of the cases, which is not significantly different from 50% (95% confidence interval: 48%, 56%). By contrast, the bank employees were substantially more dishonest in the professional identity condition (Fig. 1b). On average, they reported 58.2% successful coin flips, which is significantly above chance (95% confidence interval: 53%, 63%) and significantly higher than the success rate reported by the control group ($P = 0.033$, rank-sum test). Figure 1 shows that the treatment effect appears to be driven by two factors: (1) a higher fraction of subjects claiming the maximum earnings; and (2) an increase in incomplete cheating (that is, reporting 6, 7 and 8 successful coin flips). Assuming that subjects did not cheat to their disadvantage, the rate of misreporting is 16% in the professional identity condition. Alternatively, we can compute the fraction of subjects who cheated, which is 26% (Supplementary Methods).

Regression analysis (Extended Data Table 1) shows that the treatment effect is robust even when we control for a large set of individual characteristics such as age, gender, education, income, and nationality ($P = 0.034$, Wald test). When also controlling for business unit and experience in the banking industry, we find that employees in core business units were more dishonest than those in support units ($P = 0.008$, Wald test). However, the treatment effect is not stronger in these units because the interaction between the professional identity condition and working in a core unit is not significant ($P = 0.960$, Wald test).

There are several possible channels through which the professional identity priming increased dishonesty. One possibility is that the banking culture is characterized by norms that consider competitive behaviour to be intrinsically desirable. Perhaps the professional identity condition, together with the competitive feature in the coin tossing task, triggered this cultural aspect and, thus, increased subjects' desire to compete. To examine this, we asked subjects the following question after the priming: “How important is it to you to be the best at what you do?” However, the professional identity prime did not change subjects' desire to compete ($P = 0.861$, rank-sum test; Extended Data Fig. 1). Moreover,

regression analysis shows that the correlation between competitiveness and successful coin flips is statistically insignificant ($P = 0.642$, Wald test; Extended Data Table 3b). It is therefore unlikely that a change in subjects' desire to compete is the channel through which the treatment increased dishonesty. A related channel is that the professional identity questions might have reminded bank employees of their exposure to competitive incentive schemes and thereby increased cheating. If salience of competitive incentives were the driving force behind the increase in cheating, we should observe a stronger treatment effect in core business relative to support units, because competitive incentives are more prevalent in core units. As described earlier, however, the treatment effect in core units is similar and statistically indistinguishable from the support units.

Norm obedience is typically affected by (1) what people believe they should do, and (2) what they believe others do²⁴. In principle, the professional identity prime could have triggered beliefs that others report a higher number of successful coin flips than is true, and this could have induced subjects to report a higher number themselves. To address this, we conducted an incentivized experiment with 142 employees from the banking industry in which they had to predict other bank employees' reporting behaviour (Supplementary Discussion). If a change in beliefs was driving the treatment effect, we should observe that the professional identity questions induce bank employees to believe that other bank employees are more dishonest. However, and in contrast to this prediction, our manipulation did not affect beliefs about other bank employees' reporting behaviour ($P = 0.921$, rank-sum test; Extended Data Fig. 2).

Many critics believe that the professional culture in the banking industry promotes unethical behaviour owing to its focus on materialistic values^{9,25,26}. We asked subjects about the extent to which they endorse the statement that social status is primarily determined by financial success. It seems likely that subjects who endorse this statement are more prone to seek status through financial success, implying that their responses provide an approximation of their materialism. We found that subjects in the professional identity condition endorsed the statement significantly more strongly than those in the control condition ($P = 0.034$, rank-sum test; Fig. 2a). Moreover, a stronger endorsement of the materialistic statement is positively correlated with the reported number of successful outcomes (Spearman's $\rho = 0.237$, $P = 0.007$; Fig. 2b). These findings substantiate current concerns about the influence of materialistic values in the banking sector and indicate that the professional identity prime may have increased dishonesty through an increase in materialistic values.

Our results suggest that bank employees' compliance with the honesty norm was weakened in the professional identity condition. This raises

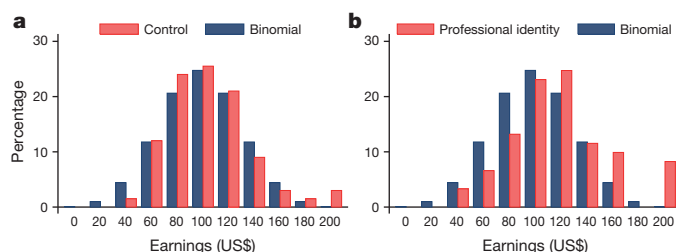


Figure 1 | Distribution of earnings in the coin tossing task claimed by the bank employees. **a, b**, Each successful coin toss yielded approximately \$20. **a**, Distribution of earnings in the control condition in comparison to the binomial distribution implied by honest reporting. On average, bank employees reported 51.6% successful coin flips, which is not significantly different from 50% ($P = 0.415$, two-sided t-test; $n = 67$). **b**, Distribution of earnings in the professional identity condition in comparison to the binomial distribution. Bank employees in the professional identity condition reported 58.2% successful coin flips, which is significantly above chance ($P = 0.002$, two-sided t-test; $n = 61$) and significantly higher than the reported success rate in the control group ($P = 0.033$, two-sided rank-sum test; $n = 128$).

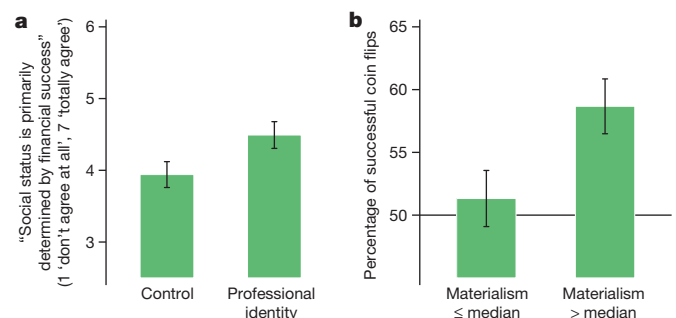


Figure 2 | Professional identity, materialism and cheating. **a**, Materialism measured by the extent to which bank employees endorse the statement “Social status is primarily determined by financial success” on a scale from 1 (not at all agree) to 7 (fully agree). Subjects in the professional identity condition show a significantly stronger endorsement of this statement ($P = 0.034$, two-sided rank-sum test; $n = 128$). **b**, Percentage of successful coin flips for participants with a below-median or median level of materialism compared to those with an above-median level of materialism. Above-median subjects are substantially more dishonest ($P = 0.010$, two-sided rank-sum test; $n = 128$). Error bars indicate standard error of the mean (s.e.m.).

the question as to whether this effect is specific to bank employees or whether it also appears in other professions. We thus repeated the experiment with 133 employees outside the banking industry. On average, they had 14.8 years of experience in a broad range of industries (for example, manufacturing, pharmaceuticals, telecommunications and information technology). Many held positions in middle or upper management. As in the case of the bank employees, the non-banking employees were randomly assigned to a professional identity or control condition (Supplementary Methods and Notes). If problematic business norms that favour dishonesty are specific to the banking industry, we should not observe a treatment effect in non-banking employees.

We found that the professional identity condition had no significant influence on dishonest behaviour in non-banking employees ($P = 0.128$, rank-sum test; Fig. 3a and Extended Data Fig. 3). A differences-in-differences regression analysis (Extended Data Table 2a) confirms that the treatment effect is significantly different for bank employees relative to non-banking employees ($P = 0.013$, Wald test), and insignificant in non-banking employees ($P = 0.191$, Wald test), suggesting that the treatment effect is specific to the business culture in the banking industry.

An alternative reason for the treatment effect in bank employees is that the banking industry is intrinsically tied to the concept of money, and that the professional identity condition rendered the concept of money salient. As there is evidence that the concept of money triggers more selfish behaviours²⁷, the professional identity condition could also have triggered more selfish, that is, dishonest, behaviour. If that were the case, it would not be the norms of the banking culture, but the salience of money that increased dishonesty in bank employees.

To examine this, we conducted another control experiment with 222 university students. Like the bank employees, the students were randomly assigned to banking or control questions. In the banking condition we asked them, for example, to name an international bank and to list tasks they think are typical for bank employees. The control questions and other aspects of the design were essentially the same as in the previous two experiments (Supplementary Methods and Notes). Thus, if an increase in money salience caused the professional identity effect,

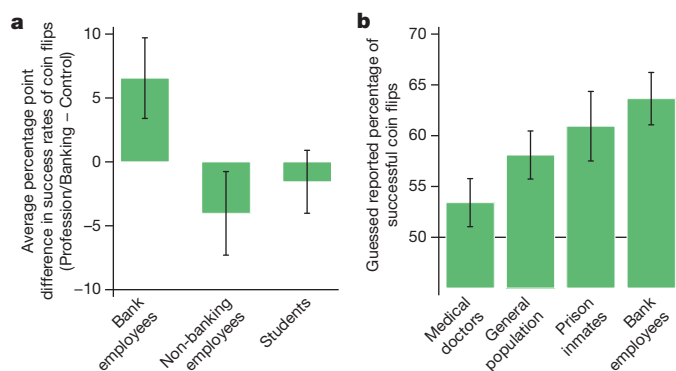


Figure 3 | Dishonest behaviour and beliefs about dishonest behaviour of different social groups. **a**, Comparison of treatment effects in the coin tossing task between bank employees, non-banking employees and students. The salience of professional identity caused an increase in dishonest behaviour in bank employees ($P = 0.033$, two-sided rank-sum test, $n = 128$). By contrast, no significant treatment effects were found in non-banking employees ($P = 0.128$, two-sided rank-sum test, $n = 133$) and students ($P = 0.390$, two-sided rank-sum test, $n = 222$). **b**, Results of a survey of individuals from the general population. Participants guessed the reported percentage of successful coin flips of either physicians ($n = 44$), bank employees ($n = 48$), prison inmates ($n = 45$) or the general population ($n = 46$). Bank employees are perceived to be less honest than physicians ($P = 0.005$, two-sided rank-sum test) and people from the general population ($P = 0.080$, two-sided rank-sum test). Bank employees are believed to behave about as dishonestly as prison inmates ($P = 0.558$, two-sided rank-sum test). Error bars indicate s.e.m.

the students in the banking condition should also behave more dishonestly than those in the control condition.

However, the data do not support the hypothesis that money salience drives the professional identity effect. The banking questions had no significant influence on dishonesty in students ($P = 0.390$, rank-sum test, Fig. 3a and Extended Data Fig. 4). A regression analysis (Extended Data Table 2b) shows that the behaviour of students and bank employees is not significantly different in the control condition ($P = 0.105$, Wald test), but—in contrast to the bank employees—the students did not change their behaviour in response to the manipulation ($P = 0.588$, Wald test). The regression analysis further confirms that the treatment effect in the bank employees and the students is significantly different ($P = 0.048$, Wald test).

Our results suggest that the prevailing business culture in the banking industry favours dishonest behaviour and thus has contributed to the loss of the industry's reputation. In contrast to their public image, however, we find that bank employees behave honestly on average in the control condition. To examine how bad bank employees' reputation is, we asked people from the general population about the percentage of successful coin flips they would expect bank employees to report in the coin tossing task. Other survey participants were asked how physicians, prison inmates, or people from the general population would behave in this task (Supplementary Methods). The participants believed that bank employees would be the most dishonest group (Fig. 3b); they believed that bank employees would report a rate of 64% successful coin flips, which corresponds to a cheating rate of 27%. This result, together with representative evidence from other sources (Extended Data Fig. 5), shows that the banking industry currently has a very bad reputation.

People's confidence in the honesty of bank employees is a key asset for the long-term stability of the financial industry. Understanding the determinants of dishonest business practices is therefore essential for the development of possible remedies. Our results suggest that banks should encourage honest behaviours by changing the norms associated with their workers' professional identity. For example, several experts and regulators have proposed that bank employees should take a professional oath analogous to the Hippocratic oath for physicians^{28,29}. Such an oath, supported by ethics training, could prompt bank employees to consider the impact of their behaviour on society rather than focusing on their own short-term benefits. A norm change also requires that companies remove financial incentives that reward employees for dishonest behaviours. In addition, existing research suggests that ethics reminders may promote compliance with the honesty norm^{4,21,30}. The use of ethics reminders requires a detailed analysis of work routines to find out where and when employees make critical decisions regarding norm obedience, so that normative demands can be rendered salient at the right time and place. These measures may be an important step towards fostering desirable and sustainable changes in business culture.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 February; accepted 17 October 2014.

Published online 19 November 2014.

- Algan, Y. & Cahuc, P. Trust and growth. *Annu. Rev. Econ.* **5**, 521–549 (2013).
- Ariely, D. *The (Honest) Truth About Dishonesty: How We Lie To Everyone—Especially Ourselves* (Harper, 2012).
- Coricelli, G., Joffily, M., Montmarquette, C. & Villeval, M. C. Cheating, emotions, and rationality: an experiment on tax evasion. *Exp. Econ.* **13**, 226–247 (2010).
- Pruckner, G. J. & Sausgruber, R. Honesty on the streets: a field study on newspaper purchasing. *J. Eur. Econ. Assoc.* **11**, 661–679 (2013).
- Kantšukov, M. & Medvedskaja, D. From dishonesty to disaster: the reasons and consequences of rogue traders' fraudulent behavior. *Adv. Ser. Manag.* **10**, 147–165 (2013).
- Abrantes-Metz, R. M., Kraten, M., Metz, A. D. & Seow, G. S. Libor manipulation? *J. Bank. Finance* **36**, 136–150 (2012).
- Sapienza, P. & Zingales, L. A trust crisis. *Int. Rev. Finance* **12**, 123–131 (2012).
- Salz, A. *The Salz Review: An Independent Review of Barclays' Business Practices* (Barclays, 2013).

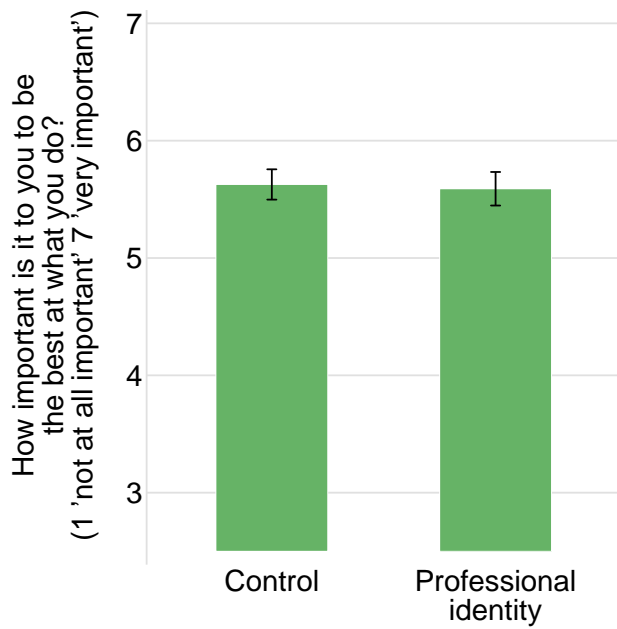
9. Hill, C. A. & Painter, R. *The Personal and Professional Responsibility of Investment Bankers* (Univ. of Minnesota Law School, 2012).
10. Moshinsky, B. *Turner Makes Claim for BOE Job as Bank Culture Reform Sought* <http://www.bloomberg.com/news/2012-07-24/turner-stakes-claim-for-boe-helm-as-bank-culture-reform-sought.html> (Bloomberg, 2012).
11. Levine, R. Financial development and economic growth: views and agenda. *J. Econ. Lit.* **35**, 688–726 (1997).
12. Hill, C. A. Bankers behaving badly? The limits of regulatory reform. *Rev. Bank. Financ. Law* **31**, 675–691 (2012).
13. Akerlof, G. A. & Kranton, R. E. *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being* (Princeton Univ. Press, 2010).
14. Benjamin, D. J., Choi, J. J. & Strickland, A. J. Social Identity and preferences. *Am. Econ. Rev.* **100**, 1913–1928 (2010).
15. Shih, M., Pittinsky, T. L. & Ambady, N. Stereotype susceptibility: identity salience and shifts in quantitative performance. *Psychol. Sci.* **10**, 80–83 (1999).
16. LeBoeuf, R. A., Shafir, E. & Bayuk, J. B. The conflicting choices of alternating selves. *Organ. Behav. Hum. Decis. Process.* **111**, 48–61 (2010).
17. Cohn, A., Maréchal, M. A. & Noll, T. *Bad Boys: How Criminal Identity Affects Rule Violation* <http://ideas.repec.org/p/zur/econwp/132.html> (Ideas, 2013).
18. Bucciol, A. & Piovesan, M. Luck or cheating? A field experiment on honesty with children. *J. Econ. Psychol.* **32**, 73–78 (2011).
19. Houser, D., Vetter, S. & Winter, J. Fairness and cheating. *Eur. Econ. Rev.* **56**, 1645–1655 (2012).
20. Fischbacher, U. & Föllmi-Heusi, F. Lies in disguise—an experimental study on cheating. *J. Eur. Econ. Assoc.* **11**, 525–547 (2013).
21. Mazar, N., Amir, O. & Ariely, D. The dishonesty of honest people: a theory of self-concept maintenance. *J. Mark. Res.* **45**, 633–644 (2008).
22. Shalvi, S., Eldar, O. & Bereby-Meyer, Y. Honesty requires time (and lack of justifications). *Psychol. Sci.* **23**, 1264–1270 (2012).
23. Ohtsubo, Y., Masuda, F., Watanabe, E. & Masuchi, A. Dishonesty invites costly third-party punishment. *Evol. Hum. Behav.* **31**, 259–264 (2010).
24. Bicchieri, C. *The Grammar of Society: The Nature and Dynamics of Social Norms* (Cambridge Univ. Press, 2006).
25. Smith, G. *Why I left Goldman Sachs: A Wall Street Story* (Grand Central Publishing, 2012).
26. Das, S. *Traders, Guns & Money: Knowns and Unknowns in the Dazzling World of Derivatives* (Financial Times/Prentice Hall, 2010).
27. Vohs, K. D., Mead, N. L. & Goode, M. R. The psychological consequences of money. *Science* **314**, 1154–1156 (2006).
28. Boatright, J. R. Swearing to be virtuous: the prospects of a banker's oath. *Rev. Soc. Econ.* **71**, 140–165 (2013).
29. Quinn, J. *HSBC Chairman Urges Bankers to Swear an Oath* <http://www.telegraph.co.uk/finance/newsbysector/banksandfinance/9700776/HSBC-Chairman-urges-bankers-to-swear-an-oath.html> (2012).
30. Shu, L. L., Mazar, N., Gino, F., Ariely, D. & Bazerman, M. H. Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proc. Natl Acad. Sci. USA* **109**, 15197–15200 (2012).

Supplementary Information is available in the online version of the paper.

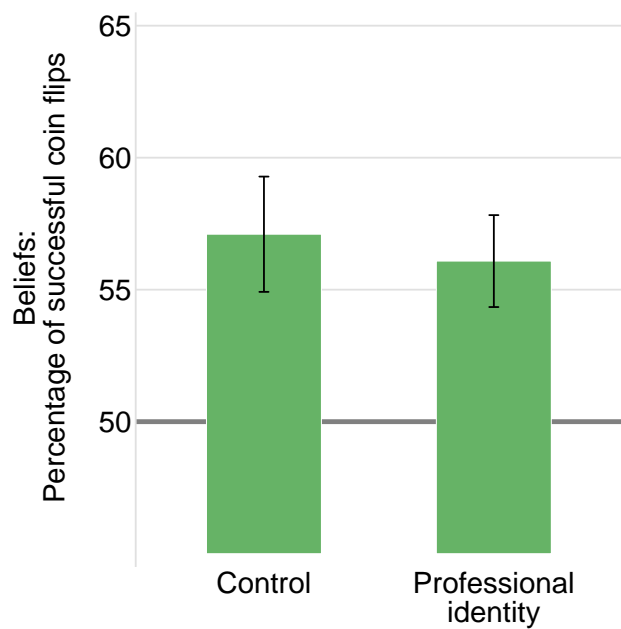
Acknowledgements We thank the participating bank, the alumni network organizations of three education programmes, and the municipal office for enabling the research; G. Akerlof and D. Ariely for critical discussions; C. Efferson, T. Herz and S. Gschwend for reading of the manuscript; and D. Bigliel and M. Brunner for research assistance. Financial support from the European Research Council Grant “Foundations of Economic Preferences” and the Gottlieb Duttweiler Institute is gratefully acknowledged.

Author Contributions A.C. and M.A.M. developed the research idea; A.C., E.F. and M.A.M. designed the study; A.C. and M.A.M. conducted the experiments, and analysed data. A.C., E.F. and M.A.M. wrote the manuscript.

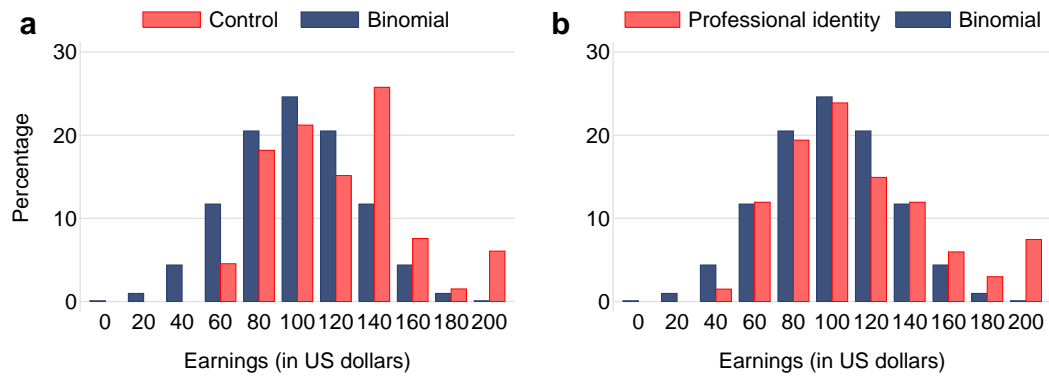
Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.C. (alain.cohn@econ.uzh.ch), E.F. (ernst.fehr@econ.uzh.ch) or M.A.M. (michel.marechal@econ.uzh.ch).



Extended Data Figure 1 | The impact of professional identity on bank employees' intrinsic competitiveness. Competitiveness is measured by the bank employees' answers to the question "How important is it to you to be the best at what you do?" on a scale from 1 (not at all important) to 7 (very important). We find no significant difference in competitiveness between the professional identity and the control condition ($P = 0.861$, two-sided rank-sum test, $n = 128$). Error bars indicate s.e.m.

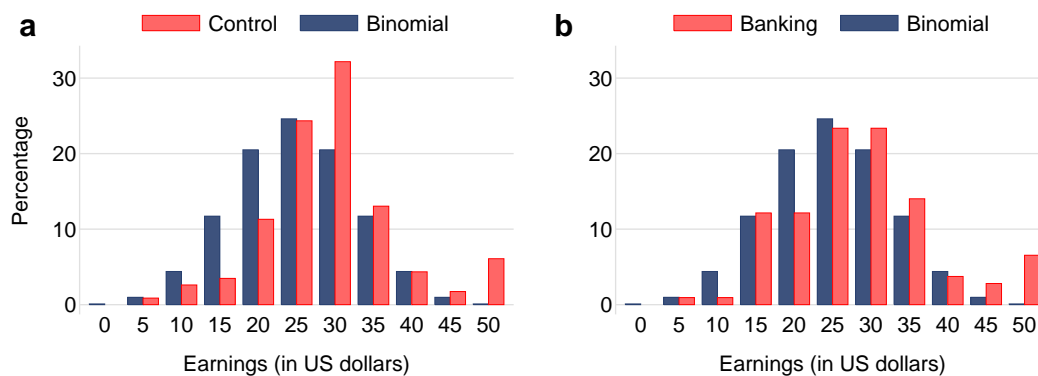


Extended Data Figure 2 | Bank employees' beliefs about other bank employees' percentage of successful coin flips. Incentivized belief elicitation experiment with bank employees. The professional identity condition had no significant influence on beliefs ($P = 0.921$, two-sided rank-sum test, $n = 142$). Error bars indicate s.e.m.



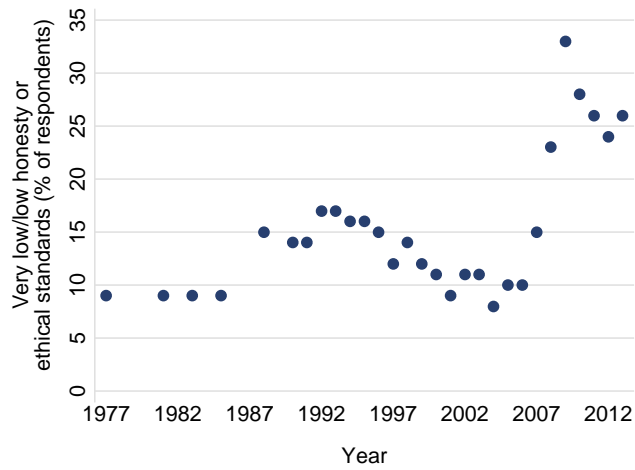
Extended Data Figure 3 | Distribution of earnings in the coin tossing task claimed by the non-banking employees. **a, b,** Each successful coin toss yielded approximately \$20. **a,** Distribution of earnings in the control condition and binomial distribution (implied by honest reporting). **b,** Distribution of earnings in the professional identity condition and binomial distribution.

The rate of successful coin flips is 55.8% (12% misreporting) in the professional identity condition, and 59.8% (20% misreporting) in the control condition, which is not significantly different ($P = 0.128$, two-sided rank-sum test, $n = 133$).



Extended Data Figure 4 | Distribution of earnings in the coin tossing task claimed by the students. **a, b,** Each successful coin toss yielded approximately \$5. **a,** Distribution of earnings in the control condition and binomial distribution (implied by honest reporting). **b,** Distribution of earnings in the

banking condition and binomial distribution. The rate of successful coin flips is 57.9% (16% misreporting) in the control condition and 56.4% (13% misreporting) in the banking condition, which is not significantly different ($P = 0.390$, two-sided rank-sum test, $n = 222$).



Extended Data Figure 5 | Gallup survey of honesty standards of people in the banking industry. Fraction of US citizens thinking that bankers have very low or low honesty or ethical standards from 1977 to 2013. This graph is an interpretation of data compiled by Gallup, Inc. However, Gallup, Inc. had no part in the creation of this graphic interpretation.

Extended Data Table 1 | Effect of professional identity on cheating in bank employees

Dependent variable:	(a)	(b)
	1 if successful outcome	
Professional identity	0.065** (0.031)	0.060** (0.030)
Age	-0.000 (0.002)	0.000 (0.003)
Male	-0.001 (0.034)	0.006 (0.033)
University degree	0.006 (0.031)	0.016 (0.034)
Foreign nationality	0.062 (0.046)	0.050 (0.049)
Relative income	-0.005 (0.012)	-0.006 (0.012)
Core business unit		0.086*** (0.033)
Years in industry		-0.000 (0.003)
Observations	1280	1280
Sample	Bank employees	Bank employees

Probit estimates. Reported results are marginal effects calculated at the median levels of the covariates. Robust standard errors corrected for clustering on the individual level are in parentheses. **a**, The decision to report a successful outcome is regressed on a dummy for the professional identity condition and individual characteristics ($n = 128$). **b**, The second model also includes job-related variables ($n = 128$). Significance levels: * $P < 0.10$, ** $P < 0.05$, *** $P < 0.01$ (two-sided Wald tests).

Extended Data Table 2 | Effect of professional identity/banking condition on cheating in bank and non-banking employees/students

Dependent variable:	(a)	(b)
	1 if successful outcome	
Professional identity/Banking condition	-0.042 (0.032)	-0.013 (0.025)
Professional identity/Banking condition X Bank employees	0.111** (0.045)	0.078** (0.040)
Bank employees	-0.082** (0.034)	-0.064 (0.040)
Age	-0.001 (0.002)	-0.000 (0.002)
Male	0.024 (0.025)	0.027 (0.020)
University degree	0.029 (0.024)	0.007 (0.030)
Foreign nationality	0.048 (0.036)	0.016 (0.026)
Relative income	-0.003 (0.009)	
Years in industry	0.002 (0.002)	
Observations	2610	3500
Sample	Bank employees and non-banking employees	Bank employees and students

Probit estimates. Reported results are marginal effects calculated at the median levels of the covariates. Robust standard errors corrected for clustering on the individual level are in parentheses. The decision to report a successful outcome is regressed on a dummy for the professional identity/banking condition and individual characteristics. We include a dummy for bank employees and an interaction term between this dummy and the treatment dummy. **a.** The first model compares the treatment effect in bank and non-banking employees ($n = 261$), which is indicated in the table by the interaction between 'Professional identity/Banking condition' and 'Bank employees'. **b.** The second model compares the treatment effect in bank employees and students ($n = 350$), which is again indicated by the interaction between 'Professional identity/Banking condition' and 'Bank employees'. We corrected the marginal effect of the interaction term by taking into account the nonlinearity of the Probit model. Significance levels: * $P < 0.10$, ** $P < 0.05$, *** $P < 0.01$ (two-sided Wald tests).

Extended Data Table 3 | Competitiveness, professional identity and cheating in bank employees

Dependent variable:	(a) 1 if successful outcome	(b)
Professional identity	0.060** (0.030)	0.060** (0.030)
Age	0.000 (0.003)	0.000 (0.003)
Male	0.006 (0.033)	0.009 (0.030)
University degree	0.016 (0.034)	0.017 (0.034)
Foreign nationality	0.050 (0.049)	0.048 (0.049)
Relative income	-0.006 (0.012)	-0.008 (0.012)
Core business unit	0.086*** (0.033)	0.087*** (0.033)
Years in industry	-0.000 (0.003)	0.000 (0.003)
Competitiveness		0.008 (0.017)
Observations	1280	1280
Sample	Bank employees	Bank employees

Probit estimates. Reported results are marginal effects calculated at the median levels of the covariates. Robust standard errors corrected for clustering on the individual level are in parentheses. **a.** The decision to report a successful outcome is regressed on a dummy for the professional identity condition and individual characteristics ($n = 128$). **b.** The second model also includes the bank employees' degree of competitiveness as an additional control variable ($n = 128$). Significance levels: * $P < 0.10$, ** $P < 0.05$, *** $P < 0.01$ (two-sided Wald tests).

Inhibition of cell expansion by rapid ABP1-mediated auxin effect on microtubules

Xu Chen^{1,2,3}, Laurie Grandont⁴, Hongjiang Li^{1,2,3}, Robert Hauschild¹, Sébastien Paque⁴, Anas Abuzeineh^{2,3}, Hana Rakusová^{1,2,3}, Eva Benkova^{1,2,3}, Catherine Perrot-Rechenmann⁴ & Jiří Friml^{1,2,3}

The prominent and evolutionarily ancient role of the plant hormone auxin is the regulation of cell expansion¹. Cell expansion requires ordered arrangement of the cytoskeleton² but molecular mechanisms underlying its regulation by signalling molecules including auxin are unknown. Here we show in the model plant *Arabidopsis thaliana* that in elongating cells exogenous application of auxin or redistribution of endogenous auxin induces very rapid microtubule re-orientation from transverse to longitudinal, coherent with the inhibition of cell expansion. This fast auxin effect requires auxin binding protein 1 (ABP1) and involves a contribution of downstream signalling components such as ROP6 GTPase, ROP-interactive protein RIC1 and the microtubule-severing protein katanin. These components are required for rapid auxin- and ABP1-mediated re-orientation of microtubules to regulate cell elongation in roots and dark-grown hypocotyls as well as asymmetric growth during gravitropic responses.

Auxin is crucial for diverse developmental processes and growth responses³. One major effect of auxin is cell expansion¹, which relies on the coordinated activities of cellular processes involving the cytoskeleton². When cells elongate, cortical microtubules are arranged perpendicularly to the axis of cell elongation (transverse microtubules), whereas a longitudinal alignment accompanies growth inhibition². The dynamic nature of microtubules provides the flexibility to rearrange them into different arrays⁴, enabling growth changes downstream of different signals such as gravity⁵ or light⁶. Many of these signalling pathways converge on auxin⁷; therefore its action upstream of microtubules translates different signals into growth responses¹. Nonetheless, whether auxin acts directly on microtubule arrangement and by which mechanism remain unclear.

Microtubules co-align approximately perpendicularly to the elongation axis in roots and dark-grown hypocotyls⁴. The transition zone of primary root and the elongation zone of etiolated hypocotyl (Extended Data Fig. 1a) are controlled by auxin to determine their respective growth rates⁷. We visualized cortical microtubules in transgenic lines expressing microtubule-associated protein 4 fused with green fluorescent protein (MAP4-GFP)⁸ or α -tubulin 6 fused with red fluorescent protein (TUA6-RFP)⁹, and classified cells on the basis of prevalent microtubule arrangement into four groups (Fig. 1a). In root, microtubules were mainly transverse and underwent visible realignment within 10 min after application of the synthetic auxin naphthalene-1-acetic acid (NAA), leading to partial longitudinal re-orientation after 1 h (Fig. 1a). Comparable effects were observed irrespective of the microtubule reporter following treatment with the natural auxin indole-3-acetic acid (IAA) (Extended Data Fig. 1b–f). The same effects were observed in etiolated hypocotyls (Fig. 1b) although at a higher auxin concentration, consistent with known auxin response maxima of aerial tissues at higher doses¹⁰. Re-orientation of microtubules is not always homogenous, as revealed by the deviated angle of individual microtubules. Transverse microtubules ($90 \pm 30^\circ$) decreased at the expense of increasingly oblique and longitudinal

microtubules ($0\text{--}60^\circ/120\text{--}180^\circ$) following auxin treatment (Extended Data Fig. 1c, f).

Treatment with the weak auxin analogue¹¹ of 2-NAA showed a very weak effect on microtubule rearrangement whereas acidic pH led to massive disruption and random orientations of microtubules (Extended

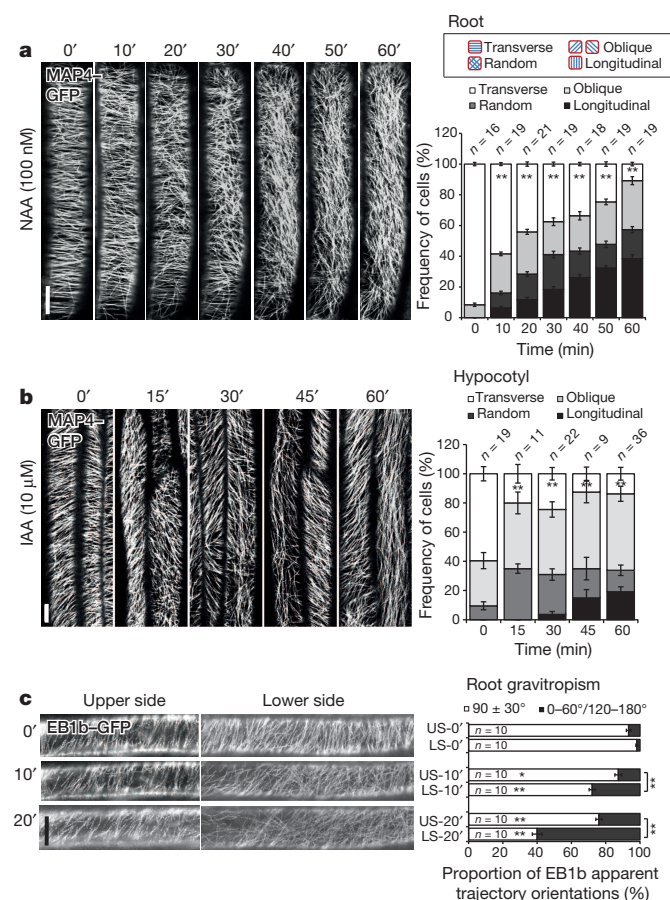


Figure 1 | Auxin induces microtubule re-orientation. **a, b**, MAP4-GFP visualization of microtubule orientation in roots (**a**) and etiolated hypocotyls (**b**) by time-lapse imaging following 100 nM NAA or 10 μ M IAA treatment, respectively. The cartoon illustrates the four categories of microtubule orientation. **c**, EB1b-GFP visualization of microtubule trajectories at the upper side (US) and lower side (LS) of 90° gravistimulated roots. EB1b trajectories were quantified as transverse ($90 \pm 30^\circ$) or longitudinal ($0\text{--}60^\circ/120\text{--}180^\circ$) microtubules. In all panels, average values are shown, error bars are s.e.m. and Student's *t*-test was calculated for transverse microtubules ($*P < 0.05$, $**P < 0.001$). Scale bars, 5 μ m (**a, c**) and 10 μ m (**b**).

¹Institute of Science and Technology Austria (IST Austria), Am Campus 1, 3400 Klosterneuburg, Austria. ²Department of Plant Systems Biology, Vlaams Instituut voor Biotechnologie (VIB), Ghent University, B-9052 Gent, Belgium. ³Department of Plant Biotechnology and Genetics, Ghent University, B-9052 Gent, Belgium. ⁴Institut des Sciences du Végétal, UPR2355 CNRS, Saclay Plant Sciences LabEx, 1 Avenue de la Terrasse, 91198 Gif sur Yvette, Cedex, France.

Data Fig. 1g, h). Both treatments confirm the specificity of active auxins on microtubule orientation.

In roots, gravistimulation induces asymmetric auxin redistribution: lower levels at the upper side correlate with cell elongation and higher levels at the lower side with inhibition of cell expansion¹². We assessed the effect of endogenous auxin redistribution on microtubule arrangement by tracking trajectories of end binding 1b (EB1b), a protein that preferentially accumulates at the growing plus ends of microtubules¹³. After 90° root re-orientation, transverse microtubules were maintained in the upper-side cells whereas at the lower side, microtubules longitudinally re-oriented within 10 min, preceding growth inhibition (Fig. 1c). Auxin distribution reported by the auxin response reporter DII-VENUS¹⁴ during gravitropism confirmed higher auxin response at the lower side than the upper side (Extended Data Fig. 1i–k). Thus auxin application or endogenous auxin redistribution promotes longitudinal microtubule orientation, correlating with auxin inhibiting cell elongation.

Next we addressed the mechanism by which auxin influences microtubule orientation. The signalling pathway of the nuclear-localized auxin co-receptors transport inhibitor response 1/auxin-related F-box (TIR1/AFB)–auxin/IAA (AUX/IAA) regulates gene transcription and mediates many plant developmental effects³. On the other hand, the ABP1 pathway regulates transcriptional¹⁵ and non-transcriptional responses such as interdigitation of pavement cells¹⁶ or clathrin-dependent endocytosis¹⁷.

In root cells of *tir1-1 abp1-1 afb2-1 afb3-1* (ref. 18), we observed normal transverse microtubules; however, they were less sensitive to auxin treatment (Extended Data Fig. 2a, b). Functional inactivation of ABP1 in the conditional SS12S and SS12K lines^{19,20} resulted in microtubule orientation defects with increasing time of ABP1 inactivation (Extended Data Fig. 2c–f), whereas *abp1-5*, harbouring a point mutation in the auxin binding pocket susceptible to impair auxin binding¹⁶, did not show altered microtubule arrangement (Fig. 2a). Nonetheless, in both cases

a severe reduction in microtubule re-orientation was observed in response to auxin (Fig. 2a). Short-term ABP1 inactivation (8 h) in dark-grown hypocotyls also impaired microtubule responsiveness to auxin without affecting the microtubule organization per se (Fig. 2b). This observation, together with a similar effect of *abp1-5* mutation, confirms that the microtubule insensitivity to auxin does not result from pre-existing microtubule alteration in these lines. Growth of ABP1-inactivated roots was previously reported to be auxin resistant²⁰, strengthening the correlation between the effect of auxin on microtubule orientation and inhibition of cell elongation.

We also explored whether ABP1 function is required for the microtubule rearrangement and differential growth response in root gravitropism. Following 90° root re-orientation, ABP1-inactivated lines showed much weaker microtubule rearrangement at the lower side than wild type (WT) (Extended Data Fig. 3a); however, gravity-induced asymmetric auxin distribution (monitored by DII-VENUS) was also less pronounced in ABP1-inactivation lines (Extended Data Fig. 3b, c compared with Extended Data Fig. 1i–k). In line with these observations, ABP1-deficient lines showed defects in gravitropic response (Extended Data Fig. 3d).

Given the mutual impact of the TIR1 and ABP1 pathways¹⁵, it is difficult to distinguish between direct and indirect effects; however, because short-term ABP1 inactivation strongly impairs auxin-mediated microtubule re-orientation, this favours the ABP1 pathway as the primary mechanism. To gain further mechanistic insights into the effect of auxin on microtubule and transcriptional compared with fast, non-transcriptional responses, we studied the kinetics of auxin's effect on microtubule re-orientations by following EB1b movement at 15 s intervals, and its trajectories. NAA did not influence the speed of EB1b movement (Extended Data Fig. 4a, b) but gradually increased oblique and longitudinal EB1b tracks (0–60°/120–180°) after 75 s (Fig. 3a and Supplementary Videos 1 and 2). To provide quantitative measures of

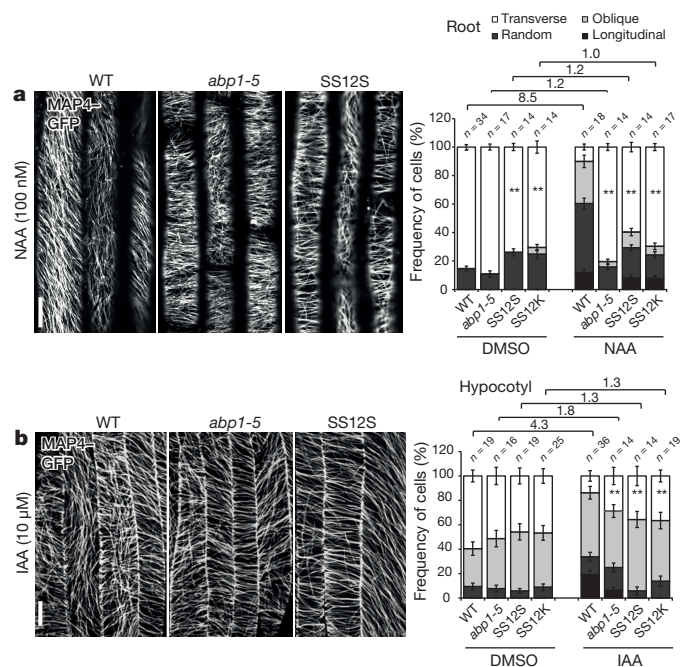


Figure 2 | ABP1 is required for auxin regulation of microtubule re-orientation. **a**, **b**, MAP4-GFP visualization of microtubule re-orientation in wild-type (WT), *abp1-5*, SS12S/K root (**a**) and hypocotyl (**b**) induced with ethanol vapours for 48 h (**a**) and 8 h (**b**) and following 60 min treatment with dimethyl sulphoxide (DMSO), 100 nM NAA (**a**) or 10 μM IAA (**b**). The ratio of transverse microtubules in DMSO versus NAA treatment is indicated above the charts (**a**, **b**). In all panels, average values are shown and error bars are s.e.m. and Student's *t*-test was calculated for transverse microtubules (***P* < 0.001). Scale bars, 5 μm (**a**) and 10 μm (**b**).

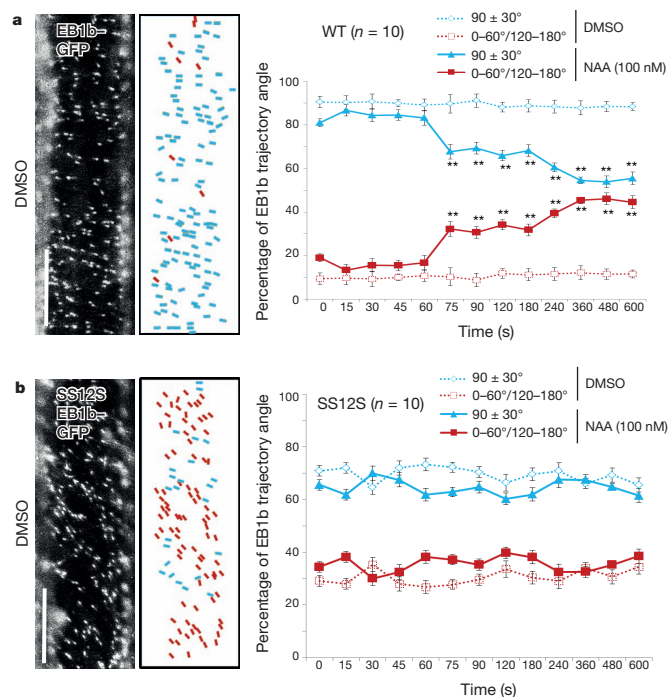


Figure 3 | The effect of auxin on fast responsiveness of microtubule rearrangement is dependent on ABP1. **a**, **b**, Projections of EB1b-GFP in WT (**a**) or SS12S (**b**) roots (left) and quantification (right) from every 15 s acquisition during 10 min (Supplementary Videos 1, 2, 3 and 5) following DMSO or 100 nM NAA application (*n* = 10). Blue and red strips represent transverse (90 ± 30°) and oblique/longitudinal (0–60°/120–180°) directions, respectively. In all panels, average values are shown and error bars are s.e.m. determined by Student's *t*-test (**P* < 0.05, ***P* < 0.001). Scale bars, 5 μm.

microtubule rearrangements, we developed an image analysis tool that automatically assigned EB1b trajectories to transverse (depicted as blue area) or longitudinal (red area) directions, which revealed changes in EB1b track orientations as soon as 30 s after auxin treatment (Extended Data Fig. 4c). Both types of kinetic analysis revealed a very fast responsiveness of microtubules to auxin, making transcriptional regulation in this process unlikely. This is consistent with the lack of interference on auxin-induced microtubule rearrangement by blocking transcription with cordycepin (Extended Data Fig. 4d). Involvement of the ABP1-dependent response pathway is also supported by the inhibition of the TIR1/AFB pathway with α -(phenylethyl-2-one)-IAA (PEO-IAA)²¹, which did not prevent the effect of IAA on microtubules (Extended Data Fig. 4e).

To investigate further the rapid ABP1-dependent effect on microtubules, we introduced EB1b-GFP in ABP1-inactivated lines. Trajectories of EB1b following ABP1 inactivation showed more oblique and longitudinal microtubules than WT (Fig. 3b, Extended Data Fig. 4f–i and Supplementary Videos 3 and 4). After NAA treatment, no consistent switch of EB1b trajectories to longitudinal directions, but only a few stochastic changes, were observed (Supplementary Videos 5 and 6, Fig. 3b and Extended Data Fig. 4f–i compared with WT in Fig. 3a and Extended Data Fig. 4c). By a complementary approach, we analysed the effect of auxin on EB1b trajectories in inducible ABP1 gain-of-function lines (XVE \gg ABP1-OE). Overexpression of ABP1-GFP in the presence of oestradiol increased the overall amount of ABP1 (Extended Data Fig. 5a–c), leading to an enhanced microtubule re-orientation in response to auxin (Extended Data Fig. 5d). In contrast, *abp1-5* exhibited delayed microtubule re-orientation in response to auxin (Extended Data Fig. 5d). Overall these results strongly suggest that the fast, non-transcriptional effect of auxin on microtubule re-orientation is mediated primarily by ABP1-dependent signalling.

Next we addressed the downstream mechanism by which ABP1 mediates the effect of auxin on microtubule arrangement. Although auxin induces calcium transients²², the manipulation of exogenous calcium had very different effects on microtubule arrangements compared with auxin (Extended Data Fig. 6). Then we tested the downstream components of the ABP1 pathway: the ROP6 GTPase, its effector RIC1 (ref. 16) and its downstream component microtubule-severing protein katanin (KTN1)^{23,24}. We analysed microtubules in *rop6-1*, *ric1-1* and *ktn1* mutants. Compared with WT, roots of *rop6-1* and *ric1-1* showed almost

normal transverse microtubules but were much less auxin responsive (Fig. 4a and Extended Data Fig. 7a). Double mutants (SS12S *ric1-1*, SS12K *ric1-1*), with ABP1 inactivation, exhibited the SS12K/S root phenotype¹⁵ but *ric1-1* microtubule arrangement (Fig. 4a and Extended Data Fig. 7a–d), consistent with the reported action of RIC1 downstream of ABP1 in early responses²⁵. The *ktn1* mutant exhibited a severe microtubule phenotype, with completely random microtubules in roots compromising further analysis. Microtubules were less disturbed in dark-grown *ktn1* hypocotyls, allowing investigation of the response to auxin and the genetic interaction with ABP1 (Extended Data Fig. 7e). Rapid auxin-induced re-orientation of microtubules was impaired in *ktn1* (Fig. 4b). Inactivation of ABP1 in *ktn1* (SS12K *ktn1*) resulted in a microtubule pattern similar to SS12K and conferred insensitivity to auxin (Fig. 4b). These data suggest that KTN1 is required for microtubule re-orientation in response to auxin but that other microtubule-associated components might be involved as well. The present data and the crucial roles of RIC1 and KTN1 in the control of microtubule architecture and crossover^{6,24} support the idea that auxin-dependent ABP1 signalling might act through Rho GTPases and RIC effectors on critical targets such as KTN1 for guiding microtubule orientation.

The effects of auxin on cell expansion are not mechanistically well understood; however, it is clear that sustainable growth control requires fast, non-transcriptional auxin effects and regulation of transcription³. Here we show that auxin signalling targets immediate changes in microtubule orientation. Auxin, by a non-transcriptional effect requiring ABP1 and downstream signalling, influences microtubule re-orientations within minutes, leading to changes in microtubules from transverse to longitudinal orientation to inhibit cell elongation. It is possible that branching of the ABP1 signalling for microtubule realignment occurs at the level of Rho GTPases, their RIC effectors and downstream targets such as KTN1. It remains unclear how this newly identified effect of ABP1 signalling on microtubules is related to other actions of the ABP1 pathway, such as inhibition of clathrin-mediated endocytosis, in particular given the differences in efficiencies of synthetic and natural auxins on both processes²⁶. The relationship between microtubules and endocytosis in plants is still unclear; however, given the observations from animals that clathrin controls microtubule acetylation²⁷, auxin-mediated defects in clathrin organization might cause the mis-modification of microtubules and thus disturb their crossover. On the other hand, Rho GTPases acting downstream of ABP1 might influence clathrin and

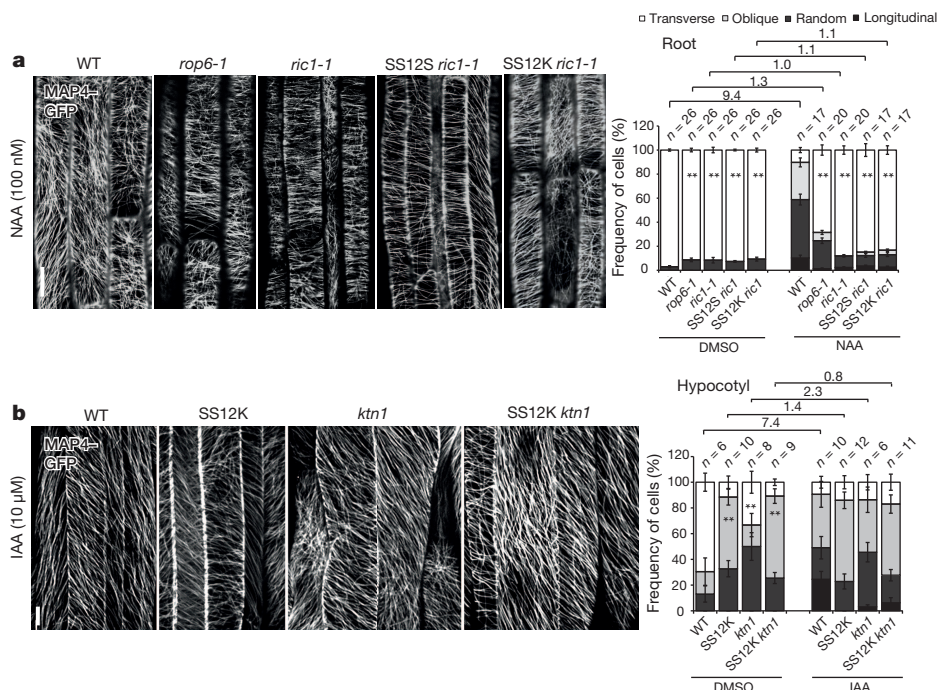


Figure 4 | Auxin-ABP1 control microtubule arrangement through downstream ROP6-RIC1 and involvement of KTN1. **a**, Microtubule orientation and quantification in roots of WT, *rop6-1*, *ric1-1*, SS12S/*ric1-1* following 60 min of DMSO or 100 nM NAA application. **b**, Microtubule orientation and quantification in 24 h ethanol-induced hypocotyls of WT, SS12K, *ktn1* and SS12K *ktn1* following 60 min of DMSO or 10 μ M IAA application. The ratio of transverse microtubules in DMSO versus NAA/IAA treatment is indicated above the charts (**a**, **b**). In all panels, average values are shown and error bars are s.e.m. and Student's *t*-test was calculated for transverse microtubules (***P* < 0.001). Scale bars, 5 μ m (**a**) and 10 μ m (**b**).

microtubule functions via distinct effectors. Our observations on the rapid regulation of microtubule arrangement by the ABP1-mediated auxin signalling pathway provide insight into the long-sought molecular mechanism by which a major plant hormone exerts its fast effect on plant cell growth.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 November 2013; accepted 23 September 2014.

Published online 17 November 2014.

- Perrot-Rechenmann, C. Cellular responses to auxin: division versus expansion. *Cold Spring Harb. Perspect. Biol.* **2**, a001446 (2010).
- Sedbrook, J. C. & Kaloriti, D. Microtubules, MAPs and plant directional cell expansion. *Trends Plant Sci.* **13**, 303–310 (2008).
- Chapman, E. J. & Estelle, M. Mechanism of auxin-regulated gene expression in plants. *Annu. Rev. Genet.* **43**, 265–285 (2009).
- Lucas, J. & Shaw, S. L. Cortical microtubule arrays in the *Arabidopsis* seedling. *Curr. Opin. Plant Biol.* **11**, 94–98 (2008).
- Blancaflor, E. B. The cytoskeleton and gravitropism in higher plants. *J. Plant Growth Regul.* **21**, 120–136 (2002).
- Lindeboom, J. J. *et al.* A mechanism for reorientation of cortical microtubule arrays driven by microtubule severing. *Science* **342**, 1245533 (2013).
- Ubeda-Tomas, S., Beemster, G. T. & Bennett, M. J. Hormonal regulation of root growth: integrating local activities into global behaviour. *Trends Plant Sci.* **17**, 326–331 (2012).
- Marc, J. *et al.* A GFP-MAP4 reporter gene for visualizing cortical microtubule rearrangements in living epidermal cells. *Plant Cell* **10**, 1927–1940 (1998).
- Chan, J., Calder, G., Fox, S. & Lloyd, C. Cortical microtubule arrays undergo rotary movements in *Arabidopsis* hypocotyl epidermal cells. *Nature Cell Biol.* **9**, 171–175 (2007).
- Collett, C. E., Harberd, N. P. & Leyser, O. Hormonal interactions in the control of *Arabidopsis* hypocotyl elongation. *Plant Physiol.* **124**, 553–562 (2000).
- Simon, S. *et al.* Defining the selectivity of processes along the auxin response chain: a study using auxin analogues. *New Phytol.* **200**, 1034–1048 (2013).
- Morita, M. T. Directional gravity sensing in gravitropism. *Annu. Rev. Plant Biol.* **61**, 705–720 (2010).
- Chan, J., Calder, G. M., Doonan, J. H. & Lloyd, C. W. EB1 reveals mobile microtubule nucleation sites in *Arabidopsis*. *Nature Cell Biol.* **5**, 967–971 (2003).
- Brunoud, G. *et al.* A novel sensor to map auxin response and distribution at high spatio-temporal resolution. *Nature* **482**, 103–106 (2012).
- Tomas, A. *et al.* Auxin-binding protein 1 is a negative regulator of the SCF^{TIR1/AFB} pathway. *Nature Commun.* **4**, 2496 (2013).
- Xu, T. *et al.* Cell surface- and rho GTPase-based auxin signaling controls cellular interdigitation in *Arabidopsis*. *Cell* **143**, 99–110 (2010).
- Robert, S. *et al.* ABP1 mediates auxin inhibition of clathrin-dependent endocytosis in *Arabidopsis*. *Cell* **143**, 111–121 (2010).
- Dharmasiri, N. *et al.* Plant development is regulated by a family of auxin receptor F box proteins. *Dev. Cell* **9**, 109–119 (2005).
- Braun, N. *et al.* Conditional repression of AUXIN BINDING PROTEIN1 reveals that it coordinates cell division and cell expansion during postembryonic shoot development in *Arabidopsis* and tobacco. *Plant Cell* **20**, 2746–2762 (2008).
- Tomas, A. *et al.* The AUXIN BINDING PROTEIN 1 is required for differential auxin responses mediating root growth. *PLoS ONE* **4**, e6648 (2009).
- Hayashi, K. *et al.* Small-molecule agonists and antagonists of F-box protein-substrate interactions in auxin perception and signaling. *Proc. Natl Acad. Sci. USA* **105**, 5632–5637 (2008).
- Shishova, M. & Lindberg, S. A new perspective on auxin perception. *J. Plant Physiol.* **167**, 417–422 (2010).
- Uyttewaal, M. *et al.* Mechanical stress acts via katanin to amplify differences in growth rate between adjacent cells in *Arabidopsis*. *Cell* **149**, 439–451 (2012).
- Lin, D. *et al.* Rho GTPase signaling activates microtubule severing to promote microtubule ordering in *Arabidopsis*. *Curr. Biol.* **23**, 290–297 (2013).
- Chen, X. *et al.* ABP1 and ROP6 GTPase signaling regulate clathrin-mediated endocytosis in *Arabidopsis* roots. *Curr. Biol.* **22**, 1326–1332 (2012).
- Paciorek, T. *et al.* Auxin inhibits endocytosis and promotes its own efflux from cells. *Nature* **435**, 1251–1256 (2005).
- Montagnac, G. *et al.* α TAT1 catalyses microtubule acetylation at clathrin-coated pits. *Nature* **502**, 567–570 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Dixit for performing complementary experiments, D. W. Ehrhardt and T. Hashimoto for providing the seeds of TUB6-RFP and EB1b-GFP respectively, E. Zazimalova, J. Petrasek and M. Fendrych for discussing the manuscript and J. Leung for text optimization. This work was supported by the European Research Council (project ERC-2011-StG-20101109-PSDP, to J.F.), ANR blanc AuxiWall project (ANR-11-BSV5-0007, to C.P.-R. and L.G.) and the Agency for Innovation by Science and Technology (IWT) (to H.R.). This work benefited from the facilities and expertise of the Imagif Cell Biology platform (<http://www.imagif.cnrs.fr>), which is supported by the Conseil Général de l'Essonne.

Author Contributions X.C., L.G., C.P.-R. and J.F. conceived the study and designed experiments. X.C. performed experiments in roots, and L.G. performed experiments in hypocotyls. H.L., S.P. and A.A. assisted in microscopy and data generation. H.R. generated partial double mutants. R.H. did bioinformatics analysis. E.B. helped with discussion of the data. X.C., L.G., C.P.-R. and J.F. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.F. (jiri.friml@ist.ac.at) or C.P.-R. (Catherine.Rechenmann@isv.cnrs-gif.fr).

METHODS

Material and growth conditions. Seeds were sown on 0.8% agar containing 1/2 Murashige and Skoog medium with sucrose for root experiments and 1% agar containing 1/2 Murashige and Skoog medium without sucrose for hypocotyl experiments. For root analyses, seedlings were grown under 16 h light/8 h dark photoperiod; for hypocotyl analyses, seedlings were grown in darkness at 22 °C for 4 days. Ethanol induction of conditional lines for ABP1 (ref. 19) (named SS12S or SS12K) was performed by exposure of the seedlings to ethanol vapour for various times as indicated for the different assays. We routinely used 48 h of exposure to vapours generated from 500 µl of 5% ethanol in light-grown root and 8 h of exposure to vapours from 500 µl of 8% ethanol for dark-grown hypocotyls (except particular induction times annotated in the figure legends). In each single experiment, WT and ABP1 conditional lines were always grown on the same plate and exposed to ethanol in identical conditions. The *ric1-1* and *rop6-1* lines were in the *Wassilewskij* (WS) background¹⁶, the other lines were derived from the *Columbia* (Col-0) background. The T-DNA insertion mutant of *KTN1* (SAIL_343_D12) was provided from the *Arabidopsis* Information Resource. Offspring of the double mutants were analysed by PCR amplification as described previously^{17,25}. For the generation of inducible overexpressing ABP1 lines (XVE >> ABP1-OE), ABP1-GFP was cloned by inserting GFP into ABP1 after the glycine 120 by primer extension PCR with two glycines flanking the GFP coding sequence¹⁷; then the fragment of full-length ABP1 genomic DNA with GFP was cloned into the Gateway vector pMDC7B using Gateway cloning technology (www.invitrogen.com). At least three independent lines were used for the analysis (H.R. et al., unpublished observations). MAP4-GFP⁸, TUA6-RFP⁹ and EB1b-GFP¹³ were used as microtubule markers, and DII-VENUS was used as an early auxin response sensor¹⁴.

Confocal microscopy observation. For observations of microtubule orientation in root, cells in the transition zone of the primary root were visualized by vertical Zeiss LSM 700 confocal laser scanning microscope (with a ×63 objective using hair gel as immersion medium). For observations in dark-grown hypocotyls, cells in the elongation zone were visualized by a Nipkow Spinning Disk confocal system (Yokogawa CXU-X1-A1) mounted on a Nikon Eclipse Ti E inverted microscope (with a ×40 oil immersion objective). For dark-grown hypocotyls, all manipulations were performed under green light to avoid any light effects on microtubule re-orientation before confocal imaging. For all the visualization of auxin-treated seedlings, 1 min manipulation time was needed to apply auxin before imaging. The videos of EB1b trajectory were performed by spinning disc confocal system with ×63 water immersion objective. Videos were acquired with 300 ms exposure time every 500 ms for 10 min. The settings of excitation and detection were as follows: for GFP, 488 nm, 505–550 nm; for VENUS, 514 nm, 527 nm; for RFP, 587 nm, 610 nm. All the images in a single experiment were captured with the same setting. In experiments where rapid microtubule re-orientation was imaged, the seedlings were gently placed in chamber slides, covered by 0.8% agar containing 1/2 Murashige and Skoog medium or placed on liquid 1/2 Murashige and Skoog medium glass slides, and then placed on a vertical Zeiss LSM 700 confocal laser scanning microscope along their original direction of growth. When gravistimulation started, slides were rotated with a rotatable stage by 90°, and epidermal cells at the upper and lower layers were individually visualized. For live imaging on DII-VENUS, immediately after the beginning of re-orientation, the seedlings were scanned every 10 min for 1 h to follow the evolution of the DII-VENUS signal. Additionally, time-lapse visualization was performed as the time indicated (single prime (′), minute; double prime (′′), second) in figures and all the experiments were repeated at least three times.

Chemical treatment. NAA (Sigma), IAA (lab 3) and PEO-IAA²¹ were dissolved in DMSO; IAA (Sigma, lab 1, 2) and 17-β-oestradiol (Sigma) were dissolved in ethanol; and cordycepin (Sigma) was dissolved in double-distilled H₂O. For oestradiol induction, 3-day-old seedlings were transferred to 2 µM 17-β-oestradiol containing 1/2 Murashige and Skoog solid medium and were grown vertically for 12 h or 48 h before observation or protein/RNA extraction. Non-induced XVE >> ABP1-OE was used as a control. For the experiments with short-term chemical treatments, 4-day-old seedlings were gently placed on chemical-containing plates according to their original direction of growth. For combined treatments with PEO-IAA or cordycepin and 1 µM IAA, 4-day-old seedlings grown vertically on 1/2 Murashige and Skoog medium were first transferred to plates containing 10 µM PEO-IAA or 400 µM cordycepin for 30 min pre-treatment (vertically grown), then transferred on the same medium plus 1 µM IAA. For the treatment with different concentrations of CaCl₂, 1/2 Murashige and Skoog medium with vitamins but without CaCl₂ was used compared with the same medium with added 1 or 10 mM CaCl₂ or with a standard 1/2 Murashige and Skoog medium (1.5 mM CaCl₂). For short-term treatments, seedlings were mounted on 0.8% agar 1/2 Murashige and Skoog medium chamber slides or liquid 1/2 Murashige and Skoog medium glass slides containing the indicated concentration of auxin, then were immediately imaged.

Gravitropic response. Four-day-old vertically grown seedlings, under light conditions, were re-oriented by 90°, and the angles deviating from the original vertical growing direction of primary roots (defined as 0°) were tracked every 30 min until 8 h.

Quantification methods. 1. To quantify the percentage of different cell types, the number of cells at a certain microtubule orientation type was calculated as a percentage of the total measured cells in each root, and different cell types were divided into transverse, oblique, random and longitudinal groups on the basis of prevalent microtubule arrangement (Fig. 1a). At least 15 different roots were analysed and at least eight cells were quantified in each root. For hypocotyls, cells were classified into a microtubule orientation group depending on the overall angles of microtubules for each cell (0–22.5°/157.5–180° for longitudinal, 22.5–67.5°/112.5–157.5° for oblique and 67.5–112.5° for transverse). Angles of microtubules were determined both manually and by using the plug-in of the ImageJ FibrilTool²⁸. At least ten different hypocotyls were analysed and 100 hypocotyl cells were quantified.

2. To quantify the oriented angles of microtubules in root cells, the number of microtubules at a certain angle range was calculated as a percentage of the total microtubules measured. The longitudinal direction parallel to and going along the growth axis was defined as 0°, the transverse direction perpendicular to the growth axis was defined as 90°, the longitudinal direction parallel to but going opposite the growth axis was defined as 180° and angles deviating from the longitudinal direction (0°) were measured. Different microtubules angles were divided into two categories: transverse direction (90 ± 30°), oblique and longitudinal direction (0–60°/120–180°). The measurement of deviated angles of microtubules was processed by ImageJ combined with Matlab. At least eight cells of different roots were randomly selected, and in each cell at least 100 microtubules were quantified.

3. For the ‘time-stack’ images of EB1b trajectory (Fig. 1c), the video of EB1b trajectory was taken every 15 s per picture in a vertical Zeiss LSM 700 confocal laser scanning microscope for 10 min in total, every 3 s per picture in a spinning disc for 10 min in total. The maximum intensity projections of all frames were stacked by ImageJ (Image > Stack > Z-project). Then, the EB1b trajectory was quantified as described previously (method 2).

4. To quantify the rate of EB1b movement, the method has previously been described in details²⁹. In brief, the manual quantification of the EB1b trajectory (Fig. 3a, b) was performed according to the directions of each single EB1b track followed with strips. As described before, the longitudinal direction parallel to and going along the growth axis was defined as 0° and the transverse direction perpendicular to the growth axis was defined as 90°. Blue strips represent the transverse direction (90 ± 30°), red strips represent the oblique and longitudinal direction (0–60°/120–180°). The deviated angle of each EB1b strip from the transverse direction was measured by ImageJ. At least ten cells of different roots were randomly selected, and in each cell at least 100 EB1b tracks were quantified.

5. Additionally, automated quantification of the directionality of EB1b trajectory used the following steps. To determine the direction of EB1b trajectory within the cells, it is necessary to eliminate the apparent motion caused by the growth of the seedling. In a first step image, stabilization was therefore performed by manually locating a prominent feature, for example the boundary between two specific cells in all frames of the video. Translating each frame accordingly then kept the prominent feature stationary. This proved more accurate than various automated image registration methods.

The temporal resolution of the videos was generally not sufficient to track individual particles over multiple frames. Thus particle tracking velocimetry resulted in a very sparse vector field. To obtain a more complete vector field that better represented the transport direction of all reporter proteins, the Horn–Schunck method of estimating optical flow was implemented and subsequently used for the analysis. In this method a regularization parameter enforces a smooth vector field which then assigns a vector of motion (speed and direction) to each pixel and frame of the video, filling in the missing flow information from the motion of the neighbouring pixels. The regularization parameter was adjusted to match the visual assessment of the flow. The resulting vector field was afterwards averaged over several frames.

The dominant direction of motion at each pixel was calculated only if the speed reached a certain threshold, which was determined by comparison with the maximum intensity projection of the video where regions with no or little protein transport appeared darker. The values chosen for the regularization parameter and the minimum speed were kept constant throughout the analysis of all videos.

The direction of motion of the remaining pixels, for which the speed was above the threshold value, was sorted into two classes: blue indicates a direction of transport of 90° ± 30°, red corresponds to 0–60°/120–180°. Finally, the areal fraction of the two classes was calculated.

6. To quantify the fluorescence, DII-VENUS fluorescence was pseudo-coloured, and the epidermal and cortical cells of similar developmental stages with symmetrical areas in roots were selected for quantification. The intensity of nuclei was extracted using the ROI tool of Fiji software (<http://fiji.sc/wiki/index.php/Fiji>), and

the sum of fluorescence in the upper- and lower-side cells of each seedling was individually quantified¹⁴.

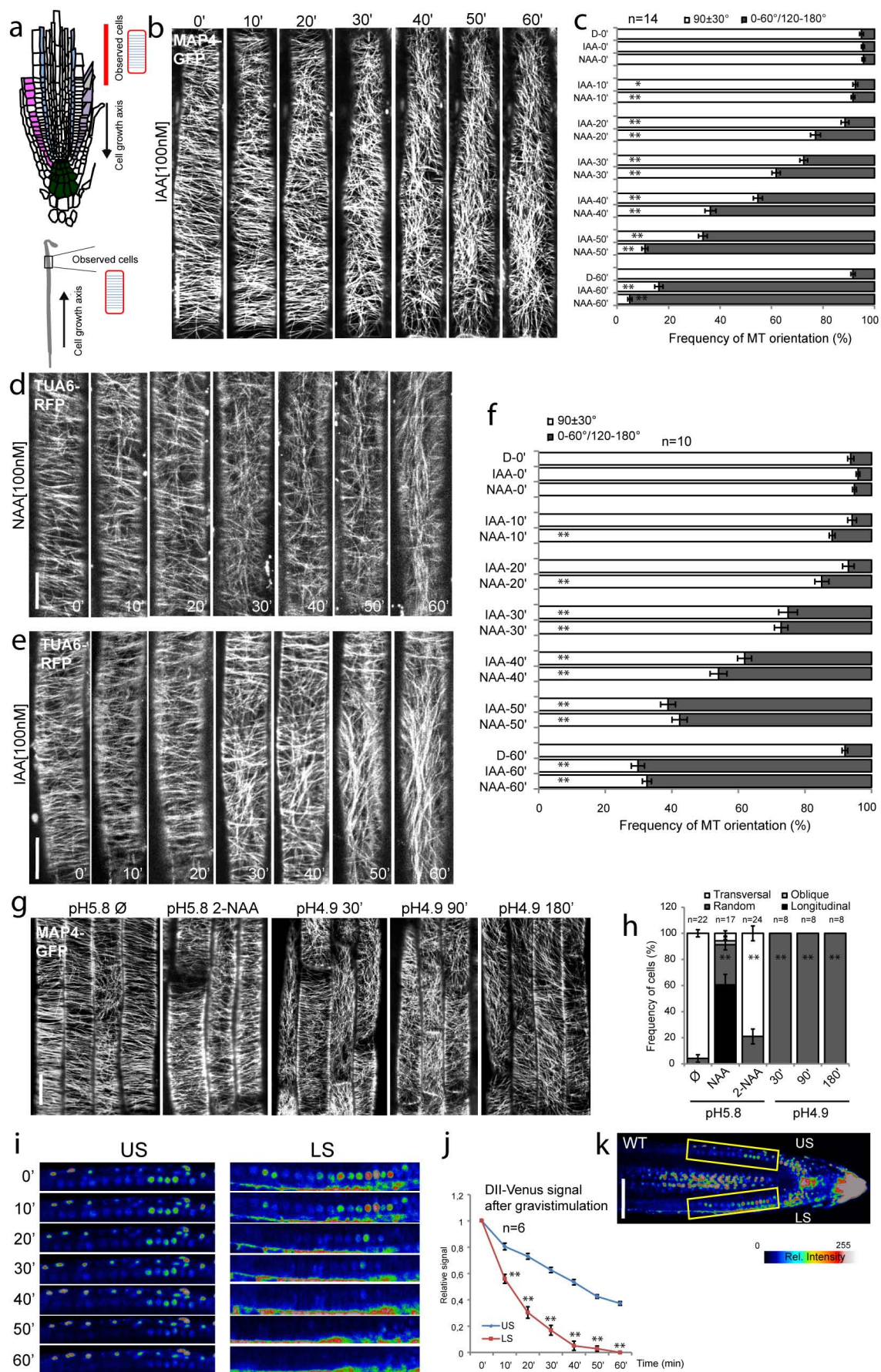
Quantitative PCR with reverse transcription. Whole RNA of seedlings was extracted using an RNeasy Mini Kit (Qiagen) and complementary DNA was synthesized using an iScript cDNA Synthesis Kit (Bio-rad). Quantitative PCR with reverse transcription (qRT-PCR) used LightCycler 480 SYBR Green I Master (Roche) following the recommendations of the manufacturer. qRT-PCR was performed in 384-well optical reaction plates by using Perkin Elmer Janus Robot and Roche Lightcycler 480 with heat for 10 min to 95 °C to activate hot-start *Taq* DNA polymerase, followed by 40 cycles of denaturation for 60 s at 95 °C and annealing extension for 60 s at 58 °C. Expression levels were normalized to the expression levels of *ACTIN8*. Specific primers of *ABP1* (F: TCGTCGTCTTTCCGTCGCG; R: TTGGCAAGCCATTGATGGGACA) and *ScFv* (F: TTAAGTGGATGCACTGGGTGA; R: AAGACTGACAGGCAGGGAGA) were used for gene expression as previously reported^{19,20}. Two biological repeats were analysed in triplicate. qRT-PCR relative quantification was performed on Lightcycler 480 software combined with the local website (<http://qpcr.ista.local>).

Protein analysis. Proteins were extracted from WT and XVE >> ABP1-OE 10-day-old seedlings induced or not with oestradiol by grinding at 4 °C with a mortar in the extraction buffer (50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 100 µM MgCl₂, 5 mM sodium ascorbate, 500 mM sucrose, phosphatase and protease inhibitors). Samples

were spun at 5,000g for 10 min at 4 °C to remove cell debris. Supernatants were centrifuged at 50,000g for 60 min at 4 °C to pellet the total membrane fraction. Pellets were re-suspended in microsomal buffer (25 mM Tris-HCl pH 6.8, 0.5 mM EDTA, 0.1 mM MgCl₂, 330 mM sucrose, 10% glycerol, protease inhibitor cocktail) for the following SDS-PAGE analysis. Protein loading was controlled by Coomassie brilliant blue staining. Protein gel blot used the mAb12 mouse monoclonal antibody³⁰ to detect ABP1 protein. Protein amount was specified by GelQuant.NET software and normalized according to sample loading. Two biological repeats were analysed in duplicate.

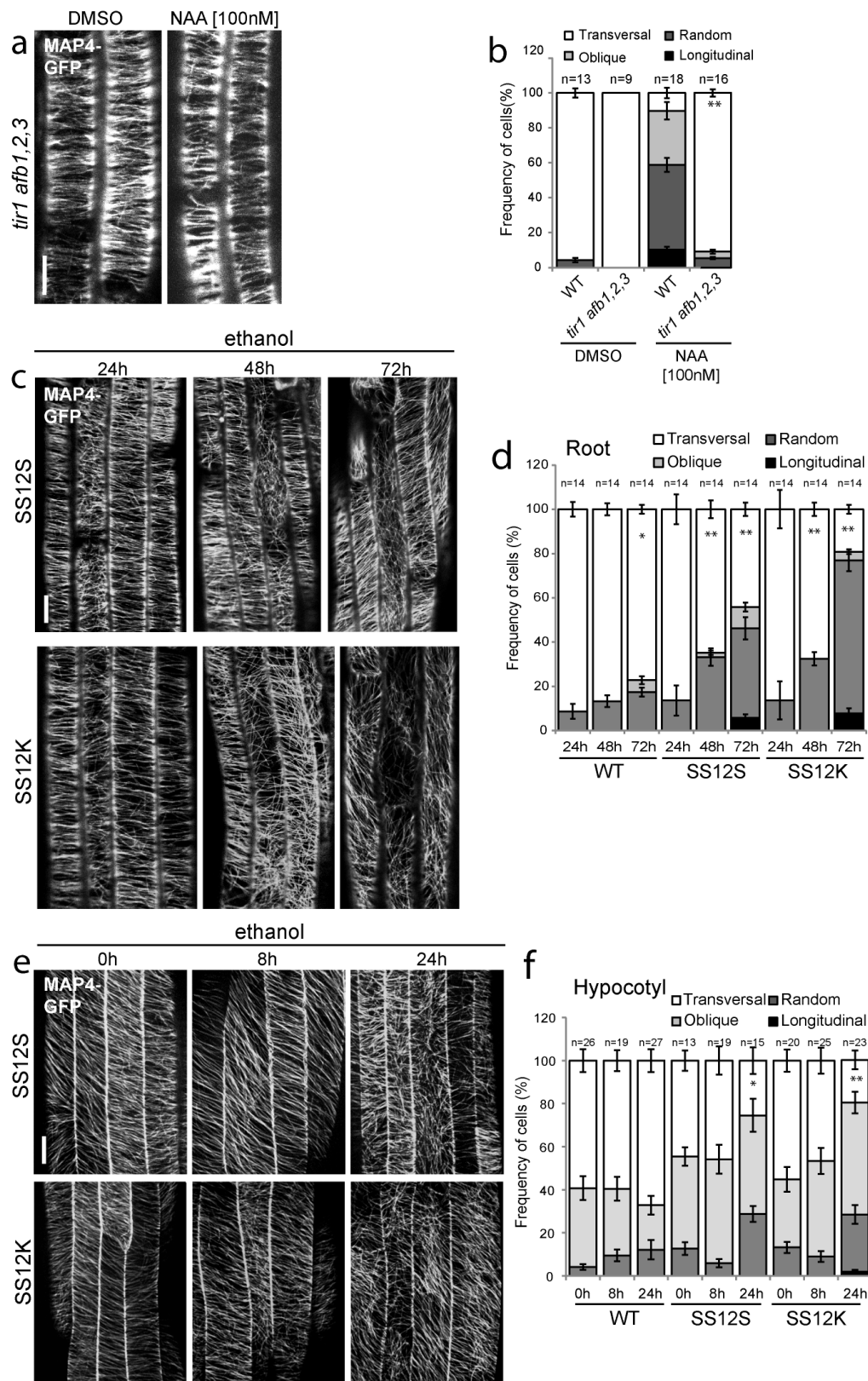
Statistics. For all quantitative data, error bars indicate s.e.m. The number of analysed samples is indicated as *n* from at least three biological replicates, and statistical analyses used Student's *t*-test where * or ** correspond to *P* < 0.05 or 0.001, respectively.

28. Boudaoud, A. *et al.* FibrilTool, an ImageJ plug-in to quantify fibrillar structures in raw microscopy images. *Nature Protocols* **9**, 457–463 (2014).
29. Salaycik, K. J., Fagerstrom, C. J., Murthy, K., Tulu, U. S. & Wadsworth, P. Quantification of microtubule nucleation, growth and dynamics in wound-edge cells. *J. Cell Sci.* **118**, 4113–4122 (2005).
30. Leblanc, N. *et al.* A novel immunological approach establishes that the auxin-binding protein, Nt-abp1, is an element involved in auxin signaling at the plasma membrane. *J. Biol. Chem.* **274**, 28314–28320 (1999).



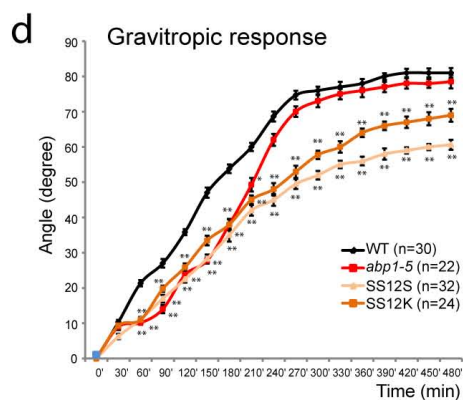
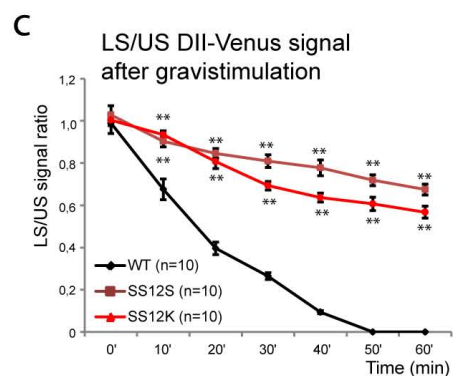
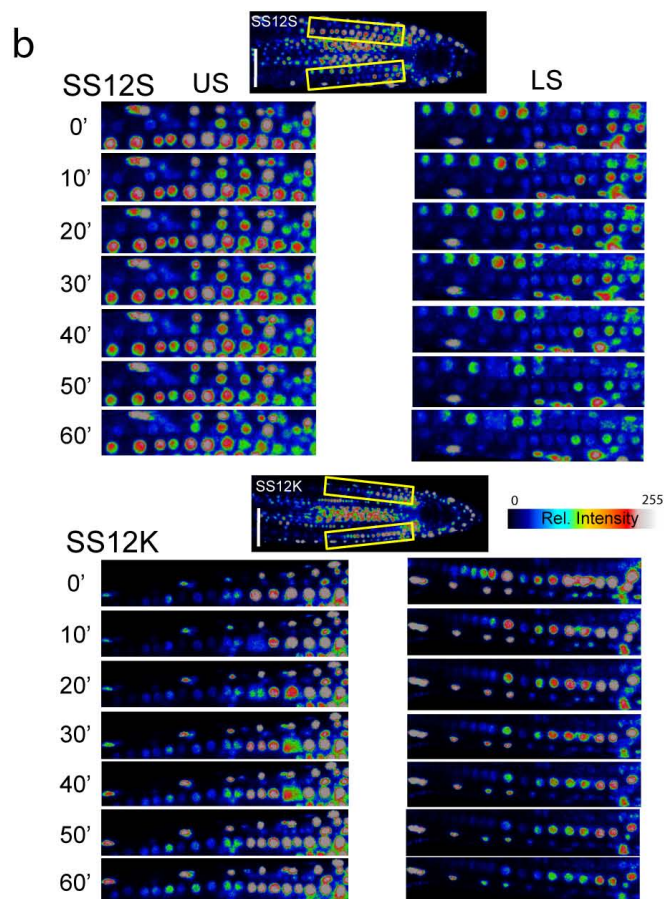
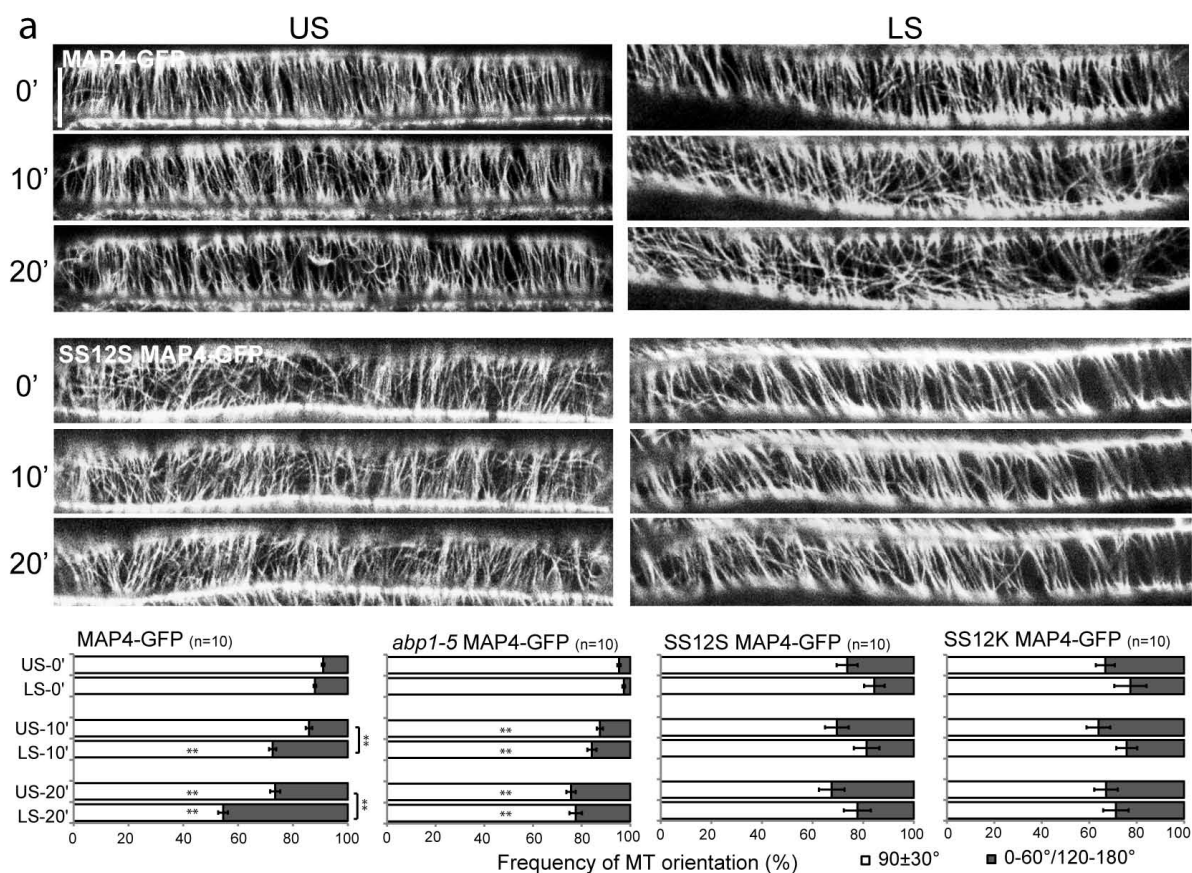
Extended Data Figure 1 | Auxin induces microtubule rearrangement in root cells. **a**, Schematic diagram of root and dark-grown hypocotyl growth. The growth direction of the root and hypocotyl is named as the cell growth axis. The observed cells for microtubules array were in the transition zone (highlighted by red line) of roots and in the elongation zone of dark-grown hypocotyls (highlighted by grey frame). The arrays of microtubules in root and hypocotyl were depicted for the expanding cells. **b–f**, MAP4–GFP or TUA6–RFP visualization of microtubule orientation in roots was performed by time-lapse observation (every 10 min; prime (′), minutes) following 100 nM NAA or IAA treatment, and deviated angles of individual microtubules were quantified as transverse microtubules ($90 \pm 30^\circ$) or longitudinal microtubules ($0\text{--}60^\circ/120\text{--}180^\circ$). In **c** and **f**, Student's *t*-test was calculated for transverse microtubules compared with untreated roots ($*P < 0.05$; $**P < 0.001$). **g, h**, MAP4–GFP visualization and quantification of microtubule orientation in

roots after 1 μM 2-NAA treatment for 60 min or after transfer of seedlings on acidified 1/2 Murashige and Skoog medium at pH 4.9 for 30, 90 and 180 minutes. Student's *t*-test was calculated for transverse microtubules in treated samples compared with 1/2 Murashige and Skoog medium (pH 5.8) growing roots used as controls ($**P < 0.001$). **i–k**, Auxin distribution approximated by DII–VENUS at the lower side (LS) and upper side (US) of 90° re-oriented WT root (in DII–VENUS background). Enlarged pictures (**i**) are shown as the frames highlighted (**k**). Signal intensity is represented by the colour code as indicated. The relative signal for the upper side and lower side (**j**) is expressed compared with the signal in the respective frame before gravistimulation. Student's *t*-test was calculated for the signal between the upper and lower sides at each time point ($**P < 0.001$). In all panels, average values are shown and error bars are s.e.m. Scale bars, 5 μm (**b, d, e, g**) and 30 μm (**k**).



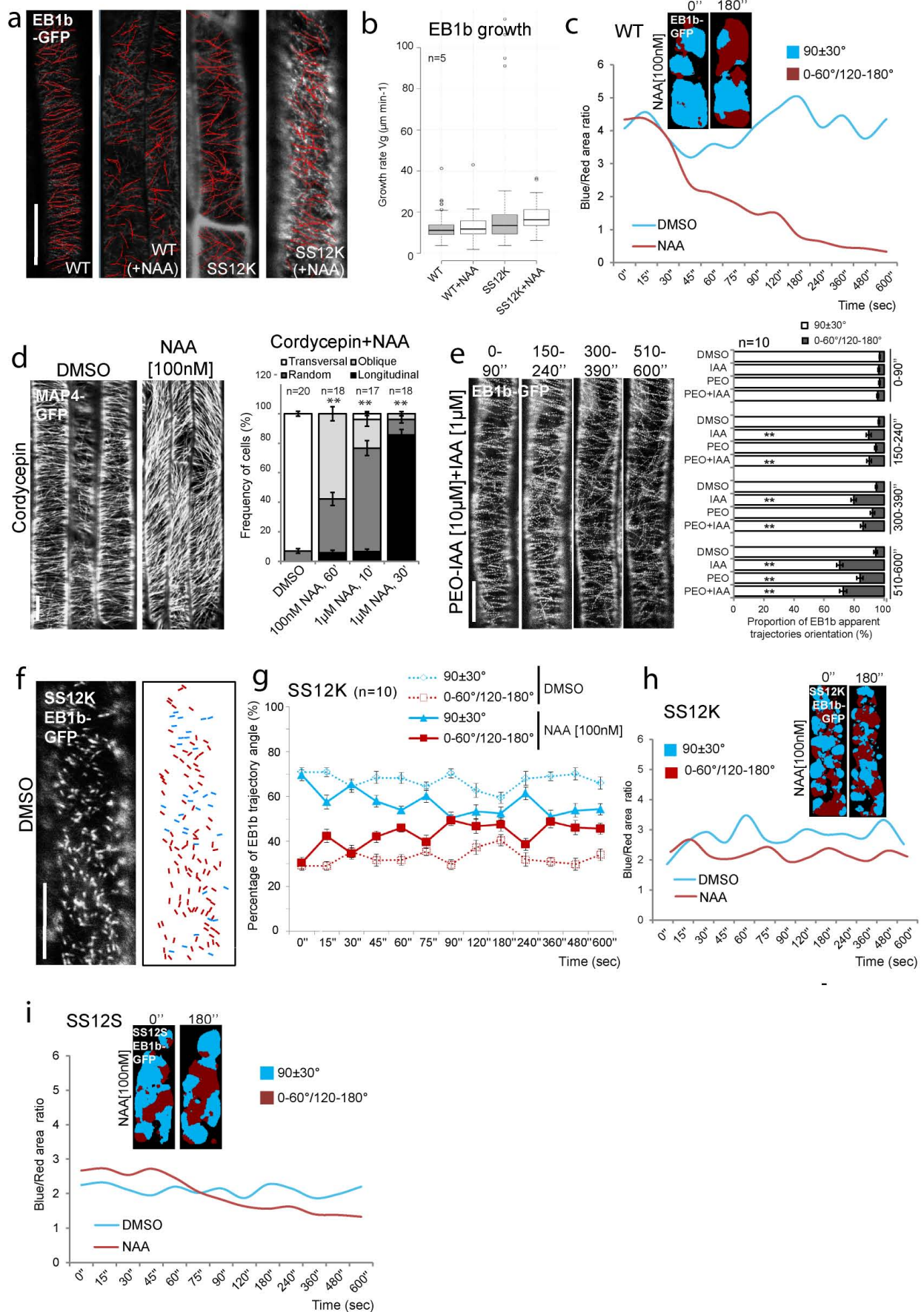
Extended Data Figure 2 | Functional inactivation of ABP1 results in microtubule defects gradually increasing with time of ABP1 inactivation. **a, b**, MAP4-GFP visualization of microtubule orientation in WT and *tir1-1afb1-1afb2-1afb3-1* (abbreviated as *tir1afb1,2,3*) seedlings following 100 nM NAA treatment for 60 min. The proportion of cells with the four categories of microtubule orientation patterns was determined, and Student's *t*-test was calculated for the category of transverse microtubules compared with WT treated in the same condition (** $P < 0.001$). **c–f**, MAP4-GFP visualization

and quantification of microtubule orientation in roots (**c, d**) or dark-grown hypocotyls (**e, f**) of WT, SS12S and SS12K seedlings following different times of ethanol induction as indicated. Student's *t*-test was calculated for the transverse microtubules compared with WT exposed for the same time to ethanol vapours as the conditional ABP1 lines (* $P < 0.05$, ** $P < 0.001$). In all panels, average values are shown and error bars are s.e.m. Scale bars, 5 µm (**a, c**) and 10 µm (**e**).



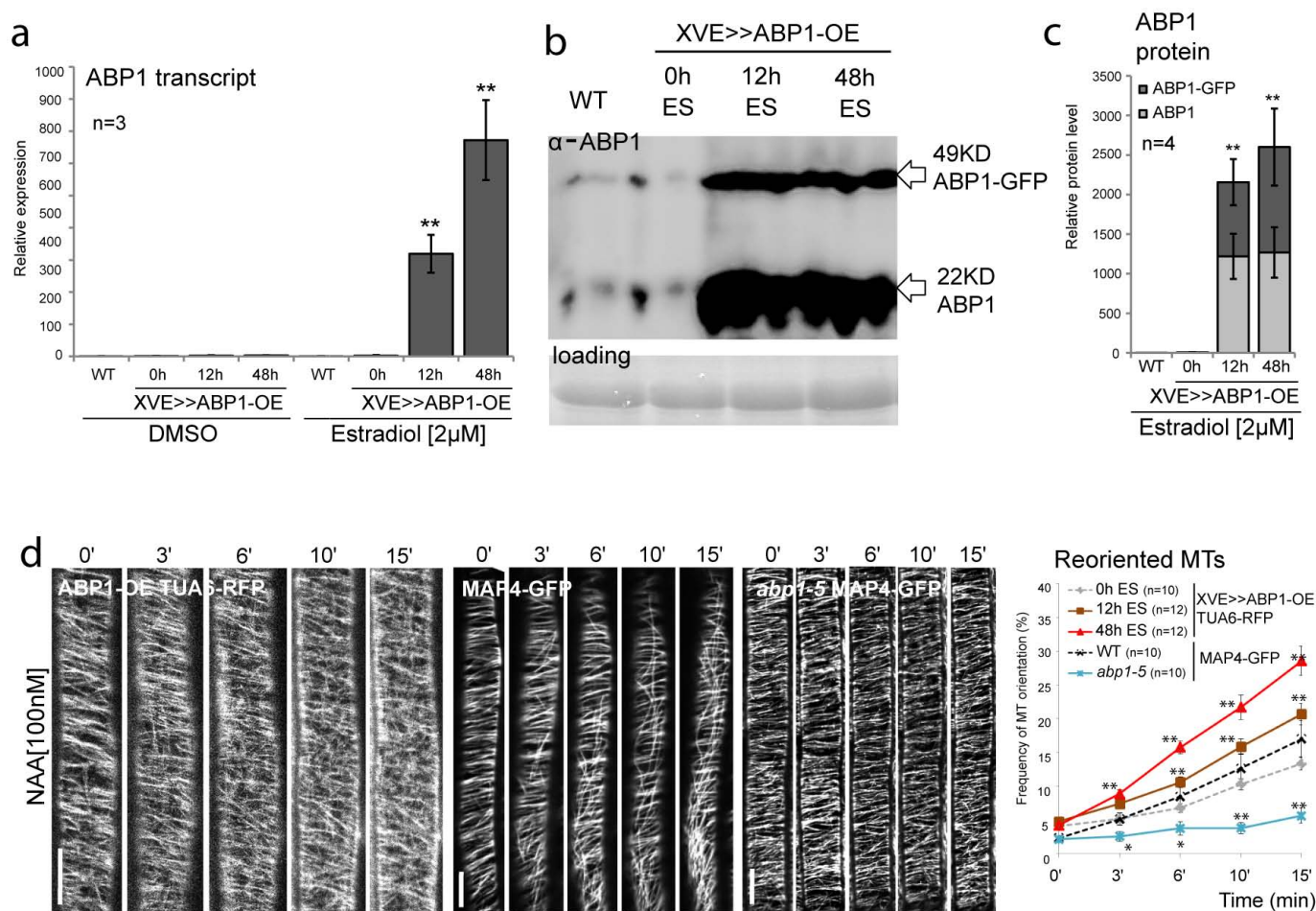
Extended Data Figure 3 | ABP1 is involved in microtubule rearrangement following gravistimulation. **a**, Rearrangement of microtubules at the lower side compared with the upper side of 90° re-oriented roots of WT, SS12S, SS12K and *abp1-5* (all expressing MAP4-GFP). Two different types of microtubule orientation ($90 \pm 30^\circ$ or $0-60^\circ/120-180^\circ$) were quantified. Student's *t*-test was calculated for the category of transverse microtubules compared with each 0' time point and calculated for transverse microtubules in the lower side compared with the upper side at each time point (** $P < 0.001$). **b**, **c**, Auxin distribution simulated by DII-VENUS at the lower side compared with the upper side of 90° re-oriented roots of SS12S and SS12K (all in DII-VENUS background; enlarged pictures were visualized in the frames

highlighted). Image stacks were taken every 10 minutes, in total for 60 minutes. The ratio of the lower side signal divided by that of the upper side is shown in the chart (**c**). Student's *t*-test was calculated for the signal ratio at each time point of SS12S/K compared with WT (** $P < 0.001$). Signal intensity is represented by the colour code as indicated. Data for SS12S and SS12K (**b**) are compared with WT (Extended Data Fig. 1i–k). **d**, The deviated angles of 90° gravistimulated-roots of WT, *abp1-5*, SS12S and SS12K seedlings were calculated every 30 min, in total for 8 h (Student's *t*-test, * $P < 0.05$, ** $P < 0.001$). In all panels, average values are shown and error bars are s.e.m. Scale bars, 5 μm (**a**) and 30 μm (**b**).



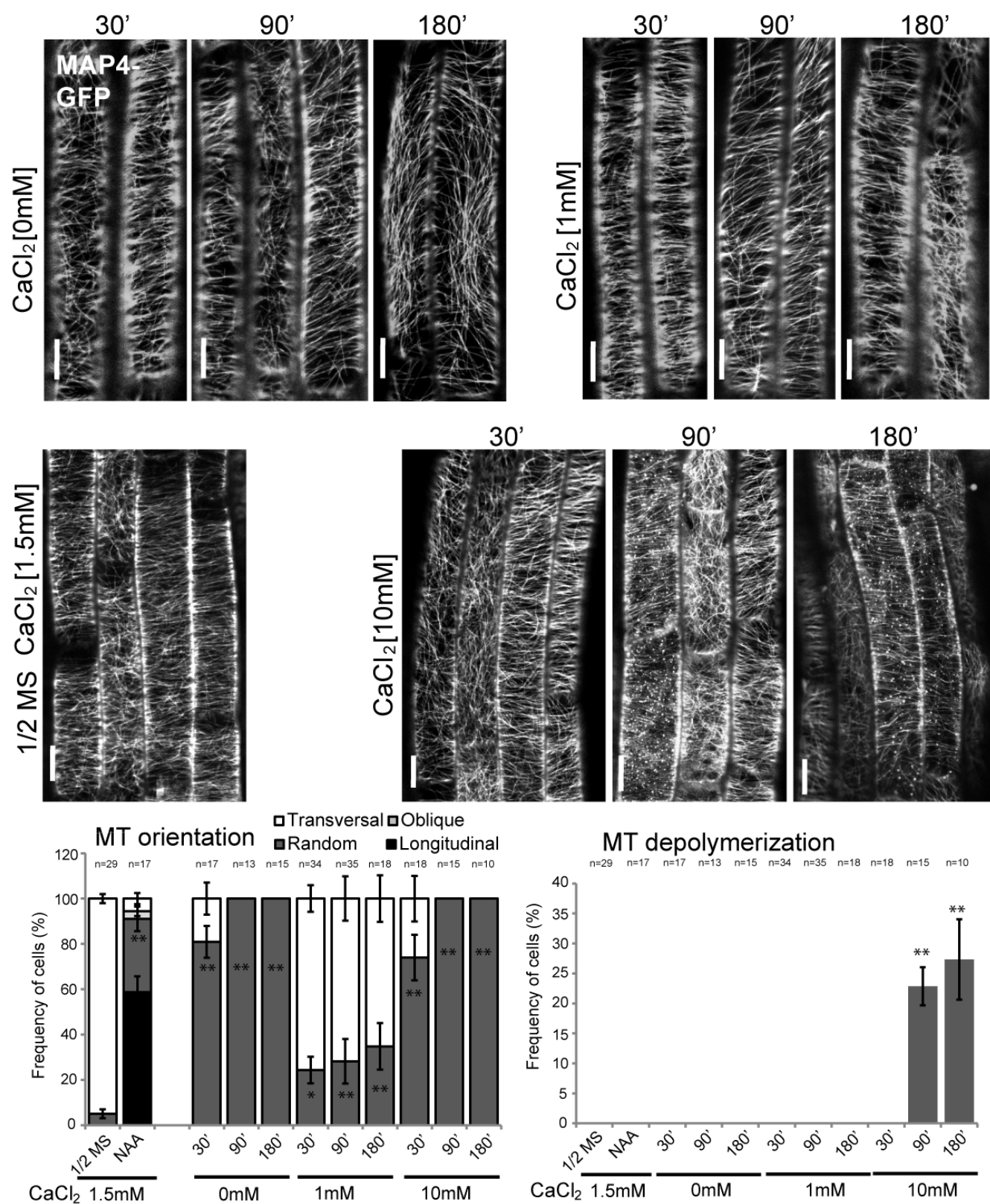
Extended Data Figure 4 | The effect of auxin on fast responsiveness of microtubule dynamics is dependent on ABP1. **a, b,** Acquisition and quantification of the rate of EB1b movement in roots of untreated or 100 nM NAA-treated (60 min) WT or SS12K (expressing EB1b-GFP) by measuring EB1b-GFP growth events as highlighted by red lines (Student's *t*-test, $P > 0.05$). Box plots indicate the 25th centile (bottom boundary), median (middle line), 75th centile (top boundary), the nearest observations within 1.5 times, the interquartile range and outliers. **c,** EB1b movement was simulated as transverse (blue, $90 \pm 30^\circ$) or longitudinal (red, $0-60^\circ/120-180^\circ$) trajectories before ($0'$) and after ($180''$) 100 nM NAA treatment in WT background (colour maps). The blue/red surface ratio is quantified on the chart ($n = 5$); **c** corresponds to Fig. 3a. **d,** Microtubule orientation patterns after 400 μ M cordycepin plus NAA co-treatment. Student's *t*-test was calculated for the category of transverse microtubule compared with only cordycepin treatment ($**P < 0.001$). **e,** EB1b trajectories (simulated by time-stack from 10 min videos) were visualized and quantified after DMSO, IAA (1 μ M), PEO-IAA

(10 μ M) and PEO-IAA (10 μ M) plus IAA (1 μ M) treatments. The left panel shows successive frames of $90'$ acquisitions following IAA application of pre-treated PEO-IAA WT roots. Student's *t*-test was calculated for the category of transverse microtubules compared with DMSO treatment at each time point ($**P < 0.001$). **f-i,** Projections of EB1b-GFP in SS12K roots (**f**) and quantification (**g**) from every 15 s acquisition during 10 min (Supplementary Videos 4 and 6) following DMSO or 100 nM NAA application ($n = 10$). Blue and red strips represent transverse ($90 \pm 30^\circ$) and oblique/longitudinal ($0-60^\circ/120-180^\circ$) directions, respectively (**f**). Colour maps show the simulated transverse or longitudinal trajectories of EB1b before ($0'$) and after ($180''$) 100 nM NAA treatment in SS12K (**h**) or SS12S (**i**) roots. The blue/red surface ratio is quantified on the charts ($n = 5$) (**h, i**). The data of SS12S (**i**) correspond to Fig. 3b, and the data of SS12S and SS12K (**f-i**) are compared with WT (Fig. 3a and Extended Data Fig. 4c). In all panels except **b**, average values are shown, error bars are s.e.m. and scale bars are 5 μ m.



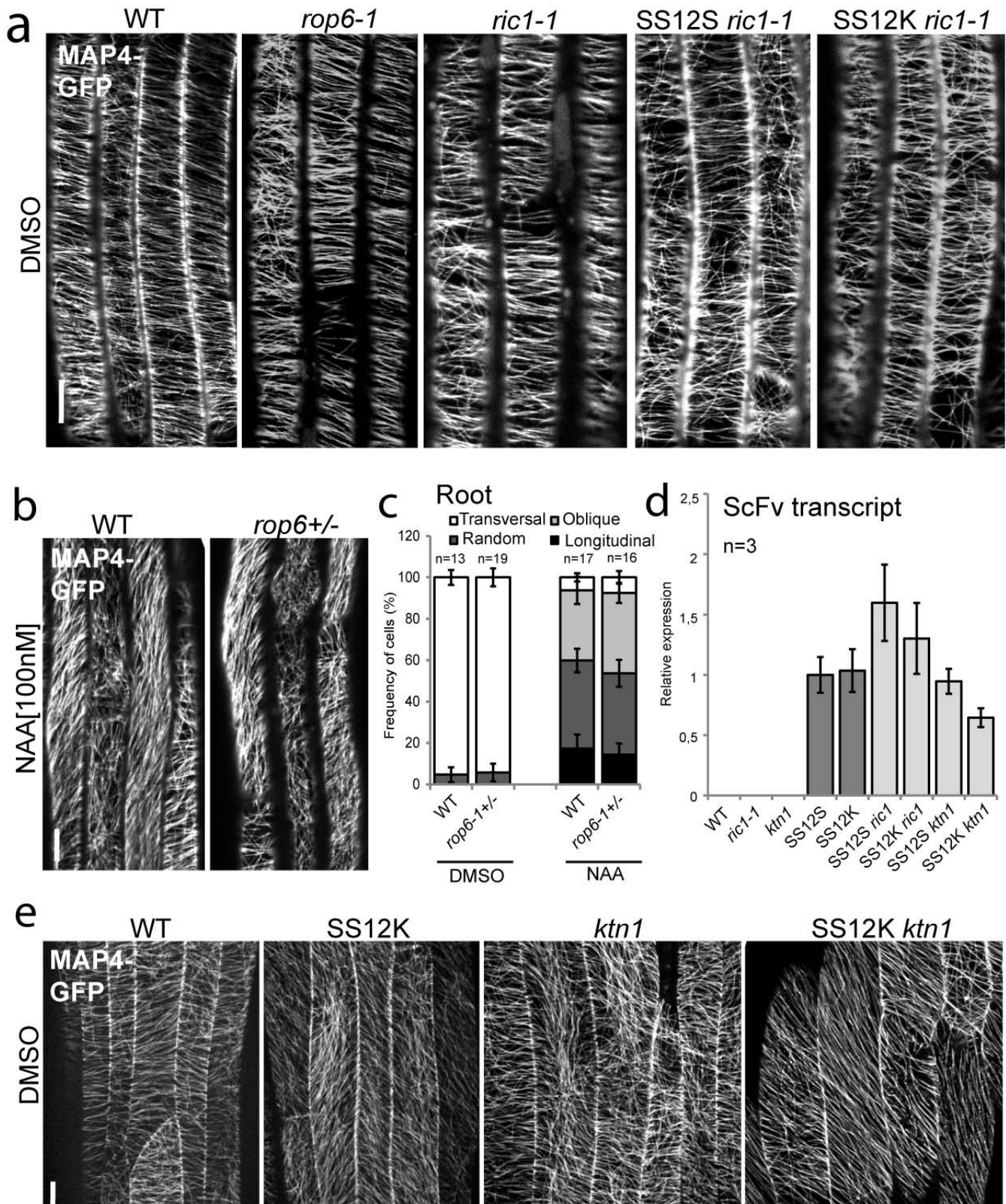
Extended Data Figure 5 | Overexpressed ABP1-induced effect of auxin on fast responsiveness of microtubule dynamics. **a–c**, ABP1 and ABP1-GFP transcripts (**a**) and ABP1 protein level (**b**, **c**) were detected in WT and XVE >> ABP1-OE line before and after 2 μM oestradiol induction for 12 h or 48 h before RNA or protein extraction. The transcript levels of ABP1 in WT with DMSO treatment were standardized as '1' (**a**). The 22 kDa native ABP1 band and 49 kDa ABP1-GFP band were detected and quantified in the right chart. The protein level of native ABP1 or ABP1-GFP in WT was standardized as '1' for each ABP1 and ABP1-GFP, respectively (**b**, **c**). Student's *t*-test,

*****P* < 0.001.** **d**, Time-lapse observation of microtubule orientation in the roots of XVE >> ABP1-OE roots expressing TUA6-RFP, WT and *abp1-5* (both expressing MAP4-GFP) upon 100 nM NAA treatment. The percentage of re-oriented microtubules (0–60°/120–180°) was quantified. Re-oriented microtubules in the inducible XVE >> ABP1-OE TUA6-RFP roots were calculated compared with none-inducible roots, and *abp1-5* MAP4-GFP was compared with MAP4-GFP at each time point (Student's *t*-test, **P* < 0.05, *****P* < 0.001**). In all panels, average values are shown, error bars are s.e.m and scale bars are 5 μm.



Extended Data Figure 6 | Calcium starvation disrupts microtubule orientation and high calcium increases microtubule depolymerization. Orientation and polymerization statuses of microtubules were visualized following transfer of seedlings to different concentrations of CaCl₂ for 30, 90 or 180 min. Low calcium levels disrupted microtubule organization, leading to a predominantly random pattern; high calcium caused microtubule

depolymerization. Student's *t*-test was calculated for the category of transverse microtubules compared with seedlings grown and transferred on standard 1/2 Murashige and Skoog medium (with 1.5 mM CaCl₂) (**P* < 0.05, ***P* < 0.001). In all panels, average values are shown, error bars are s.e.m. and scale bars are 5 μm.



Extended Data Figure 7 | Auxin-ABP1 controls microtubule arrangement through downstream ROP6-RIC1-KTN1 signalling. **a**, MAP4-GFP visualization of microtubule orientation in the root of WT, *rop6-1*, *ric1-1*, SS12S *ric1-1* and SS12K *ric1-1* following DMSO application for 60 min. Pictures in **a** correspond to quantifications in Fig. 4a. **b**, **c**, Microtubule re-orientation patterns were visualized by MAP4-GFP in the roots of WT and *rop6-1*^{+/-} following DMSO or 100 nM NAA application for 60 min (Student's *t*-test, *P* > 0.05). **d**, Transcript level of the *scFv12* coding the recombinant antibody

responsible for ABP1 knockdown in WT, *ric1-1*, *ktn1*, SS12S, SS12K, SS12S *ric1-1*, SS12K *ric1-1*, SS12S *ktn1* and SS12K *ktn1* after 48 h ethanol induction. The transcript level of the *scFv12* in SS12S was standardized as '1' (Student's *t*-test, *P* > 0.05). **e**, Microtubule orientation by MAP4-GFP in dark-grown hypocotyls of WT, SS12K, *ktn1* and SS12K *ktn1* (with 24 h ethanol induction) following DMSO application for 60 min. Pictures in **e** correspond to Fig. 4b. In all panels, average values are shown and error bars are s.e.m. Scale bars, 5 μ m (**a**, **b**) and 10 μ m (**e**).

An enteric virus can replace the beneficial function of commensal bacteria

Elisabeth Kernbauer^{1,2}, Yi Ding^{3,4} & Ken Cadwell^{1,2}

Intestinal microbial communities have profound effects on host physiology¹. Whereas the symbiotic contribution of commensal bacteria is well established, the role of eukaryotic viruses that are present in the gastrointestinal tract under homeostatic conditions is undefined^{2,3}. Here we demonstrate that a common enteric RNA virus can replace the beneficial function of commensal bacteria in the intestine. Murine norovirus (MNV) infection of germ-free or antibiotic-treated mice restored intestinal morphology and lymphocyte function without inducing overt inflammation and disease. The presence of MNV also suppressed an expansion of group 2 innate lymphoid cells observed in the absence of bacteria, and induced transcriptional changes in the intestine associated with immune development and type I interferon (IFN) signalling. Consistent with this observation, the IFN- α receptor was essential for the ability of MNV to compensate for bacterial depletion. Importantly, MNV infection offset the deleterious effect of treatment with antibiotics in models of intestinal injury and pathogenic bacterial infection. These data indicate that eukaryotic viruses have the capacity to support intestinal homeostasis and shape mucosal immunity, similarly to commensal bacteria.

Despite significant limitations in the ability to detect and annotate eukaryotic viruses present in the gastrointestinal tract, recent deep-sequencing efforts reveal the existence of a complex enteric virome^{3,4}. Members of this viral component of the intestinal microbiota could include pathogens such as noroviruses that continue to persist after disease is resolved⁵, Anelloviridae and Circoviridae family members that are ubiquitously detected in healthy individuals^{6–8}, and uncharacterized viruses that display little sequence identity with known viruses^{9–11}. Also, examination of the enteric virome of rhesus monkeys infected with simian immunodeficiency virus (SIV) suggest that many viruses are present at low levels and are kept in check by the immune system¹². Evidence that such viruses in the intestine can contribute to physiology beyond acute diarrhoeal disease is provided by studies examining MNV, a positive-strand RNA virus of the Caliciviridae family that is endemic in mouse facilities¹³. MNV displays tropism for myeloid cells and can establish persistent infection without causing obvious disease in immuno-competent mice^{13,14}. We recently demonstrated that persistent infection by the MNV strain CR6 (MNV.CR6) induces intestinal pathologies in mice deficient in the inflammatory bowel disease gene *Atg16l1* (ref. 15). The observation that MNV induces inflammatory pathologies in a genetically susceptible host resembles similar observations made with commensal bacteria. However, the bacterial component of the microbiota also provides considerable benefit to the host by generating metabolites, promoting the development of the mucosal immune system, and preventing colonization by pathogenic microorganisms¹. It is unknown whether eukaryotic viruses in the intestine interact with the host in an analogous symbiotic manner.

Much of our knowledge on the role of commensal bacteria in mucosal immunity comes from the characterization of germ-free (GF) mice, which show aberrant intestinal morphology and deficiencies in the lymphocyte compartment due to the absence of bacteria¹⁶. Thus, we used GF

mice as a reductionist model to determine whether MNV infection can provide developmental cues that have been mainly attributed to bacteria. GF mice were mono-associated with MNV.CR6 (GF+MNV) in a gnotobiotic isolator by infecting breeding pairs with 3×10^6 plaque-forming units (p.f.u.) and allowing the virus to be naturally propagated to offspring in subsequent generations, analogous to newborns that inherit commensal bacteria from parents (Extended Data Fig. 1a). Despite persistent presence of the virus, there were no signs of overt inflammation in the intestine (Extended Data Fig. 1b–d).

While GF mice had thin villi containing few CD3⁺ T cells and narrow crypts, the general appearance of the small intestine in GF+MNV mice resembled that of conventional mice (Fig. 1a–f). The presence of virus in GF mice also partially restored the number of granules and lysozyme expression in Paneth cells (Extended Data Fig. 2a–d). These changes in appearance due to MNV.CR6 were associated with a significant increase in the overall cellularity of the lamina propria and mesenteric lymph nodes (MLNs) (Fig. 1g, h). Consistent with the ability of certain bacterial species to promote lymphocyte differentiation¹, GF mice had reductions in the numbers of CD4⁺ T cells in the lamina propria and MLNs, CD8⁺ T cells in MLNs, IFN- γ expression by these cells, and mucosal and serum antibody production (Extended Data Fig. 2e–p). By contrast, GF+MNV mice displayed increases in these factors, in some cases to an extent similar to or in excess of conventional mice (Extended Data Fig. 2e–p). These effects of the virus were not dependent on neonatal infection since a reversal of abnormalities was observed when adult GF mice were infected with MNV.CR6 for 10 days (Extended Data Fig. 3). Also, the effects of MNV.CR6 cannot be explained by uncontrolled viral replication because GF mice harbour less virus compared with conventional mice upon infection (Extended Data Fig. 1b), as seen with other intestinal viruses^{17,18}. Importantly, these changes in response to MNV.CR6 are specific to conditions in which bacteria are absent. MNV.CR6 had little or no effect on the appearance of the intestine and lymphocytes in conventional mice (Extended Data Fig. 4). Therefore, there is considerable overlap between the responses to MNV and to commensal bacteria.

The host response evoked by MNV could depend on anatomical location or on the strain being examined. Abnormalities in the morphology of the colon of GF mice were not as apparent as in the small intestine, but we detected a significant increase in natural killer (NK) T cells¹⁹. While MNV mono-association had little effect on tissue morphology or on the number of NK T cells, the presence of the virus increased the frequencies of both CD4⁺ and CD8⁺ T cells in the colon (Extended Data Fig. 5). To determine whether the effect of MNV is specific to the CR6 strain, we infected GF mice with MNV.CW3, a strain that is considered more relevant to human noroviruses because it displays increased virulence compared with MNV.CR6 and establishes an acute infection²⁰. Unlike conventional mice, which clear the virus between days 5–7 after infection²⁰, we found that a low amount of MNV.CW3 could be detected in the stool of GF mice on day 10 after infection (Extended Data Fig. 1b), and that the virus strain was able to induce similar effects on intestinal morphology and the lymphocyte compartment as MNV.CR6 (Extended Data

¹Kimmel Center for Biology and Medicine at the Skirball Institute, New York University School of Medicine, New York, New York 10016, USA. ²Department of Microbiology, New York University School of Medicine, New York, New York 10016, USA. ³New York Presbyterian Hospital, New York, New York 10065, USA. ⁴Department of Pathology, New York University School of Medicine, New York, New York 10016, USA.

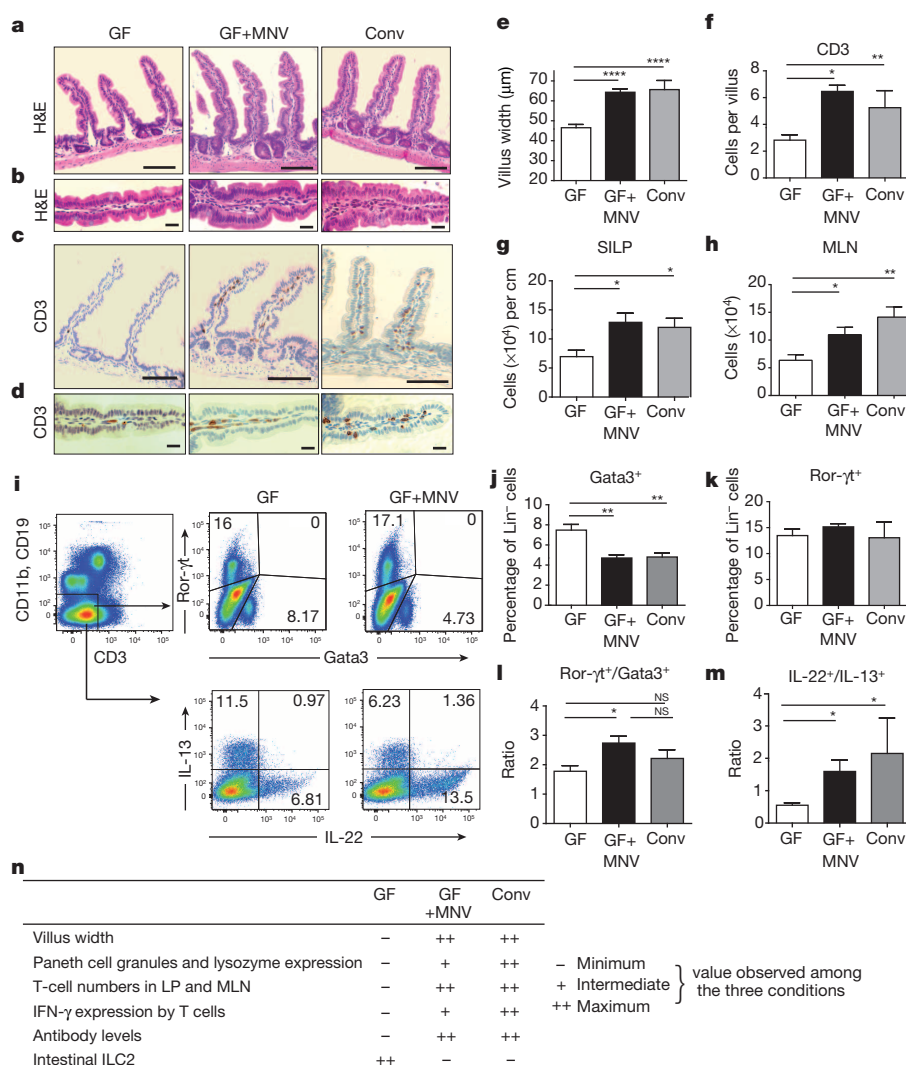


Figure 1 | MNV reverses intestinal abnormalities in GF mice.

a–d, Representative small intestinal sections from GF, GF+MNV or conventional (Conv) mice stained with haematoxylin and eosin (H&E) (**a**, **b**) or anti-CD3 antibody (**c**, **d**). Scale bars, 100 μm (**a**, **d**); 10 μm (**c**, **d**).

e, **f**, Quantification of villus width (**e**) and CD3 $^+$ cells (**f**) per villus. Fifty villi were quantified from 6–8 mice per group. **g**, **h**, Total number of cells in the small intestine lamina propria (SILP) and MLNs. $n = 7$ –11 mice per group. **i**, Flow cytometry analysis of small intestine lamina propria cells for Gata3, Ror- γ^t , IL-22 and IL-13 expression in live, Lin $^-$ cells (CD11b $^-$, CD19 $^-$ and CD3 $^-$). **j–m**, Per cent Gata3 $^+$ (**j**) and Ror- γ^t (**k**) cells, and ratio of Ror- γ^t to Gata3 $^+$ (**l**) and IL-22 $^+$ to IL-13 $^+$ (**m**) in Lin $^-$ cells from **i**. $n = 10$ mice per

group. **n**, Summary of comparisons between GF, GF+MNV and Conv mice. GF+MNV mice have similar villus width, T-cell numbers in small intestine lamina propria cells and MLNs, and antibody levels compared with conventional mice, and are designated with double plus signs, indicating restoration to the maximal value. GF+MNV mice show partial increases in the number of Paneth cell granules, lysozyme expression and IFN- γ expression. For ILC2 cells, GF mice display the maximal value and GF+MNV mice are similar to conventional mice. NS, not significant. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$. Analysis of variance (ANOVA) with Holm–Sidak correction. Graphs show means \pm standard error of the mean (s.e.m.) from at least two independent experiments.

Fig. 6b–h). Next, we isolated virus directly from a naturally infected mouse housed in the vivarium (MNV.SKI, described in Methods). Alignment of the capsid sequence showed that MNV.SKI is a previously uncharacterized strain (Extended Data Fig. 6a). MNV.SKI was also able to reverse abnormalities in GF mice (Extended Data Fig. 6b–h). Although all three MNV strains induce qualitatively similar alterations in intestinal morphology and lymphocytes, the effects are quantitatively distinct, especially when comparing T-cell subsets. Nevertheless, the general ability of MNV to evoke a response in GF mice appears to be strain independent.

Innate lymphoid cells (ILCs) are emerging as important participants in mucosal immunity, but the role of the microbiota in maintaining this compartment is less understood. Unlike the skin and lungs, the intestine has been shown to contain reduced numbers of Gata3 $^+$ ILC type 2 (ILC2) cells relative to Ror- γ^t ILC3 cells²¹. We found that GF mice had an increase in the proportion of ILC2 cells in the small intestine compared with conventional mice (Fig. 1i–m). This expansion of ILC2 cells

due to the absence of bacteria was reversed in GF+MNV mice (Fig. 1i–m). Consistent with this finding, MNV.CR6 infection increased the proportion of ILCs expressing interleukin (IL)-22 relative to IL-13, effector cytokines expressed by ILC3 cells and ILC2 cells, respectively (Fig. 1i, m). The proportion of the T-bet $^+$ ILC1 subset was similar across conditions (Extended Data Fig. 2q). Thus, we find that commensal bacteria prevent expansion of intestinal ILC2 cells, and MNV.CR6 infection mimics this function of commensal bacteria. When taken together, these results indicate that the presence of a single virus can reverse many of the abnormalities that arise in the complete absence of the commensal bacterial community (Fig. 1n).

GF mice lack bacteria from birth and represent an extreme condition. We found that adult mice treated with a cocktail of antibiotics for 2 weeks display several of the abnormalities described earlier in GF mice, including aberrant intestinal morphology and a reduction in the number of total CD4 $^+$ T cells as well as IFN- γ -expressing CD4 $^+$ and CD8 $^+$ T cells

(Fig. 2a–j and Extended Data Fig. 7a–d). Thus, we examined the ability of MNV infection to reverse these abnormalities induced by antibiotics treatment as another assay to demonstrate functional redundancies between this virus and commensal bacteria (Fig. 2a and Extended Data Fig. 1b). By day 10 after infection, MNV.CR6 infection had led to increased villus width, Paneth cell granules, intestinal T cells, and IFN- γ expression by CD4 $^{+}$ T cells (Fig. 2c–m). For comparison, mice were treated with antibiotics for 2 weeks and then inoculated with 1×10^7 colony-forming units (c.f.u.) of the representative commensals *Bacteroides thetaiotaomicron*, *Lactobacillus johnsonii*, or left alone after cessation of antibiotic treatment (Extended Data Fig. 7e–h). *B. thetaiotaomicron* and *L. johnsonii* restored the morphology of the intestine to a similar degree to MNV.CR6, but had varying effects on the T-cell compartment. *B. thetaiotaomicron* increased T-cell numbers when measured by CD3 staining of small intestinal sections, whereas *L. johnsonii* induced an IFN- γ response in CD4 $^{+}$ and CD8 $^{+}$ T cells without affecting their overall numbers (Extended Data Fig. 7i–p). These results are consistent with the concept that individual members of the microbiota evoke overlapping responses from the host, but are not necessarily identical.

To further elucidate the response evoked by MNV and commensal bacteria, we performed RNA deep-sequencing analyses (RNA-seq) on whole intestinal tissue harvested from untreated GF mice, GF mice 10 days after inoculation with MNV.CR6, and GF mice reconstituted with the bacterial flora from conventional mice (Fig. 3a, Extended Data Table 1 and Supplementary Table 1). Pathway analysis of the transcripts induced by commensal bacteria indicated alterations in the metabolic state, immunity and blood vessel morphogenesis. MNV.CR6 induced a more limited

gene expression pattern associated with lymphoid cell development and immune responses, many of which are related to the antiviral type I IFN (IFN-I) response (Extended Data Table 1a). Consistent with the shared ability of MNV and bacteria to reverse abnormalities in GF mice, the overlapping gene set contained genes associated with the development and function of haematopoietic lineage cells (Fig. 3a and Extended Data Table 1c). The IFN-I signature induced by MNV.CR6 prompted us to examine whether the effect of viral infection was dependent on IFN-I signalling. The intestine of untreated IFN- α receptor knockout mice (*Ifnar1* $^{-/-}$) displayed a normal appearance, while antibiotics treatment of these mice led to similar abnormalities in intestinal morphology and T-cell numbers as described earlier (Fig. 3b–h). However, MNV.CR6

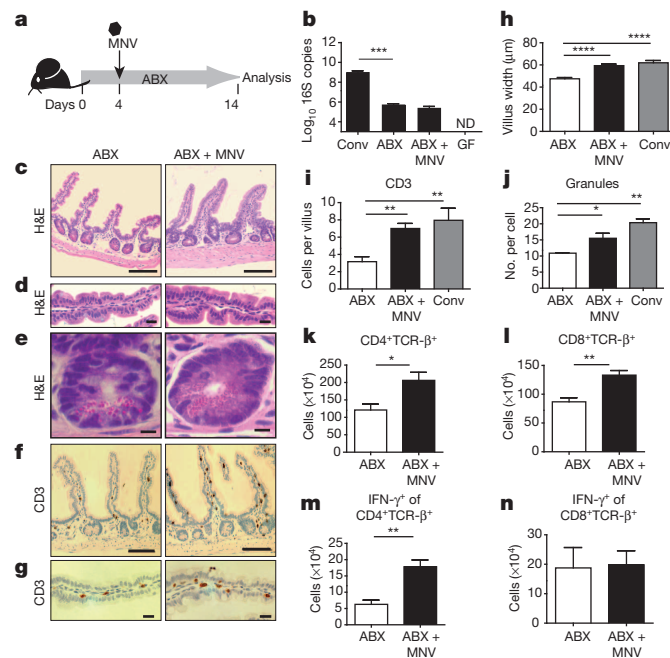


Figure 2 | MNV reverses intestinal abnormalities induced by antibiotics. **a**, Schematic for introducing MNV.CR6 into antibiotics (ABX)-treated mice (ABX+MNV). **b**, 16S quantitative reverse transcription–polymerase chain reaction (qRT–PCR) analysis of stool from conventional (Conv), antibiotics-only (ABX), antibiotics plus MNV and GF mice. ND, not detected. **c–g**, Representative small intestinal sections from antibiotics-only and antibiotics plus MNV mice stained with H&E or anti-CD3 antibody. Scale bars, 100 μ m (**c**, **f**); 10 μ m (**d**, **g**); 1 μ m (**e**). **h–j**, Quantification of villus width (**h**), CD3 $^{+}$ cells per villus (**i**) and granules per Paneth cell (**j**) for the mice in **c–g**. $n = 6–10$ mice per group. **k**, **l**, Number of CD4 $^{+}$ (**k**) and CD8 $^{+}$ T (TCR- β) $^{+}$ cells (**l**) in small intestine lamina propria cells of antibiotics-only and antibiotics plus MNV mice. **m**, **n**, Number of IFN- γ $^{+}$ CD4 $^{+}$ (**m**) and CD8 $^{+}$ (**n**) T cells for mice in **k** and **l**. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. **b**, **h**, **i**, **j**, ANOVA with Holm–Sidak correction. **k–n**, Unpaired two-tailed t test. Graphs show means \pm s.e.m. from at least two independent experiments.

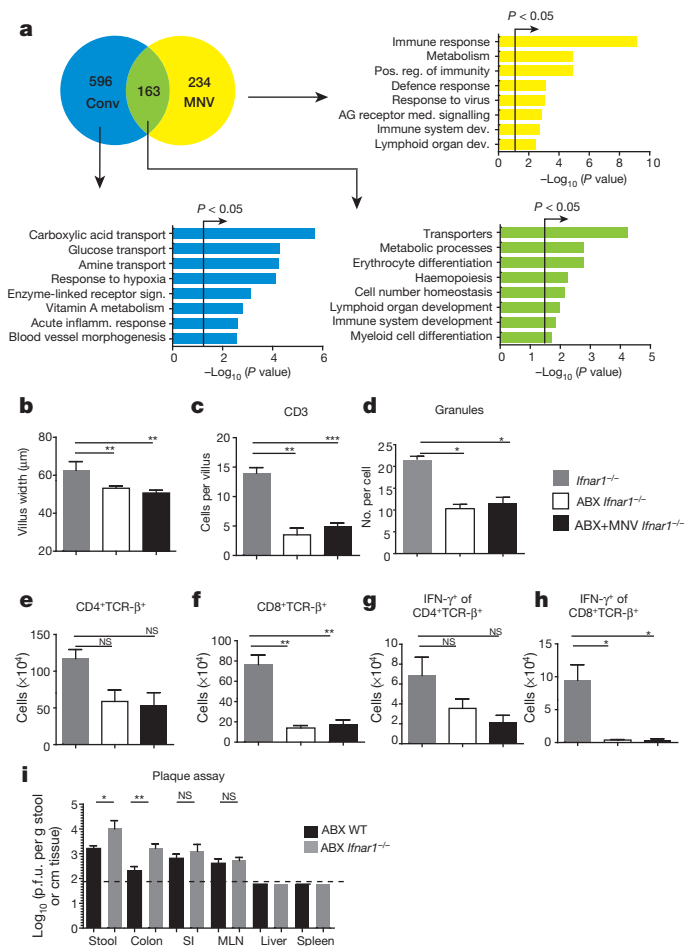


Figure 3 | MNV-induced changes are dependent on a type I IFN response. **a**, RNA-seq analysis of small intestine tissue from GF mice inoculated with MNV.CR6, conventionalized with bacteria (Conv), or left untreated. Venn diagram represents number of transcripts displaying >1.4 fold enrichment upon conventionalization (blue), inoculation with MNV.CR6 (yellow), or in both conditions (green), compared with untreated mice. Bar graphs represent Gene Ontology (GO) terms displaying association with the above gene sets. $n = 3–4$ mice. AG, antigen; dev., development; inflamm., inflammatory; med., mediated; pos., positive; reg., regulation. **b–d**, Quantification of villus width (**b**), CD3 $^{+}$ cells per villus (**c**) and granules per Paneth cells (**d**) of *Ifnar1* $^{-/-}$ mice that received antibiotics (ABX), antibiotics plus MNV.CR6 (ABX+MNV), or that were left untreated. Fifty villi or 30 crypts per mouse were quantified. $n = 6$ mice per group. **e–h**, Total number of CD4 $^{+}$ (**e**) and CD8 $^{+}$ (**f**) T cells, and IFN- γ $^{+}$ CD4 $^{+}$ (**g**) and CD8 $^{+}$ (**h**) T cells in the small intestine lamina propria cells of *Ifnar1* $^{-/-}$ mice that received antibiotics, antibiotics plus MNV.CR6, or that were left untreated. **i**, MNV.CR6 p.f.u. per cm tissue or per g stool from antibiotic-treated wild-type (WT) and *Ifnar1* $^{-/-}$ mice 10 days after infection. Dashed line denotes limit of detection. $n = 3$ mice per group. NS, not significant. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. ANOVA with Holm–Sidak correction. Graphs show means \pm s.e.m.

infection did not have a detectable effect on antibiotic-treated *Ifnar1*^{-/-} mice despite productive replication (Fig. 3i and Extended Data Fig. 1b), indicating that virus-mediated reversal of these abnormalities is dependent on IFN-I signalling. Inducing IFN-I gene expression through polyinosinic:polycytidylic acid (poly(I:C)) injection was sufficient to increase villus width in GF mice but did not have a noticeable effect on intestinal T cells (Extended Data Fig. 8). Thus, IFN-I signalling probably functions in conjunction with other pathways to confer the full effect of MNV infection. Indeed, transcriptional analyses indicate a large overlap between

the genes induced by MNV.CR6 and by bacterial colonization of GF mice that are not associated with IFN-I (Extended Data Table 1c).

The depletion of commensal bacteria is considered a major health hazard of indiscriminate use of antibiotics. Because MNV can reverse abnormalities observed upon bacterial depletion, we tested the possibility that viral infection can provide protection against intestinal damage. We found that antibiotic treatment increased sensitivity to chemical injury of the intestine by dextran sodium sulphate (DSS) (Fig. 4a). Remarkably, all three MNV strains examined were able to improve survival in antibiotic-treated mice receiving DSS (Fig. 4a). MNV.CR6, in particular, conferred robust protection and prevented shortening of the colon (Fig. 4a, b). MNV.CR6 was unable to improve survival in antibiotic-treated *Ifnar1*^{-/-} mice, indicating that IFN-I signalling was necessary (Fig. 4c). Next, we used inflammation induced by *Citrobacter rodentium* to test the protective effect of MNV, because antibiotic pre-treatment was shown to exacerbate the disease induced by this model enteric pathogen without altering bacterial burden²². In mice pre-treated with antibiotics, MNV.CR6 did not alter the amount of *C. rodentium*, which was the dominant bacterium in the gut (Fig. 4f and Extended Data Fig. 9a, b). However, MNV.CR6 ameliorated weight loss, diarrhoea and histopathology (Fig. 4d–k). *C. rodentium* induces expression of virulence factors in response to host immunity and competition with other bacteria²³. Interestingly, MNV infection of antibiotic-treated mice restored expression of bacterial virulence factors to levels detected in conventional mice (Extended Data Fig. 9c, d). This effect of the virus on bacterial gene expression and the protection from inflammation are probably due to the ability of MNV.CR6 to enhance the immune response to *C. rodentium*, which includes increased levels of antibodies as well as antimicrobial and regulatory cytokines (Extended Data Fig. 9e–j). Thus, MNV enhances protection against both infectious and non-infectious secondary challenges to the gastrointestinal tract in antibiotic-treated mice.

RNA viruses are routinely detected in asymptomatic infants and children as well as individuals recovering from acute gastroenteritis^{5,24,25}. Thus, in addition to periods after the administration of antibiotics, the presence of known or unidentified viruses may be beneficial during developmental or recovery stages when the immune system and the intestinal microbiota are vulnerable. However, like bacteria, whether a given intestinal virus is a mutualist is likely to depend on host factors and the specific viral strain in question. For instance, production of IFN-I is a conserved response to viral infections, but the role of this cytokine in the gut is highly contextual and dependent on other soluble factors²⁶. For this reason, it may not be surprising that MNV can protect antibiotic-treated mice from intestinal injury, but exacerbates disease in mice that are genetically susceptible to inflammatory bowel disease^{15,27}. We speculate that viruses that contribute to the pathogenesis of complex inflammatory disorders could be commensal in nature rather than exclusively pathogenic. Taken together, we demonstrate that an enteric animal virus can function in a manner analogous to commensal bacteria, and propose that symbiotic viruses exist in the mammalian gastrointestinal tract. We suggest that the incredible advances in viral discovery pipelines²⁸ should be leveraged to discover the commensal or mutualistic viruses that constitute the virome, in addition to traditional pathogens.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 June; accepted 13 October 2014.

Published online 19 November 2014.

1. Honda, K. & Littman, D. R. The microbiome in infectious disease and inflammation. *Annu. Rev. Immunol.* **30**, 759–795 (2012).
2. Duerkop, B. A. & Hooper, L. V. Resident viruses and their interactions with the immune system. *Nature Immunol.* **14**, 654–659 (2013).
3. Virgin, H. W. The virome in mammalian physiology and disease. *Cell* **157**, 142–150 (2014).
4. Norman, J. M., Handley, S. A. & Virgin, H. W. Kingdom-agnostic metagenomics and the importance of complete characterization of enteric microbial communities. *Gastroenterology* **146**, 1459–1469 (2014).

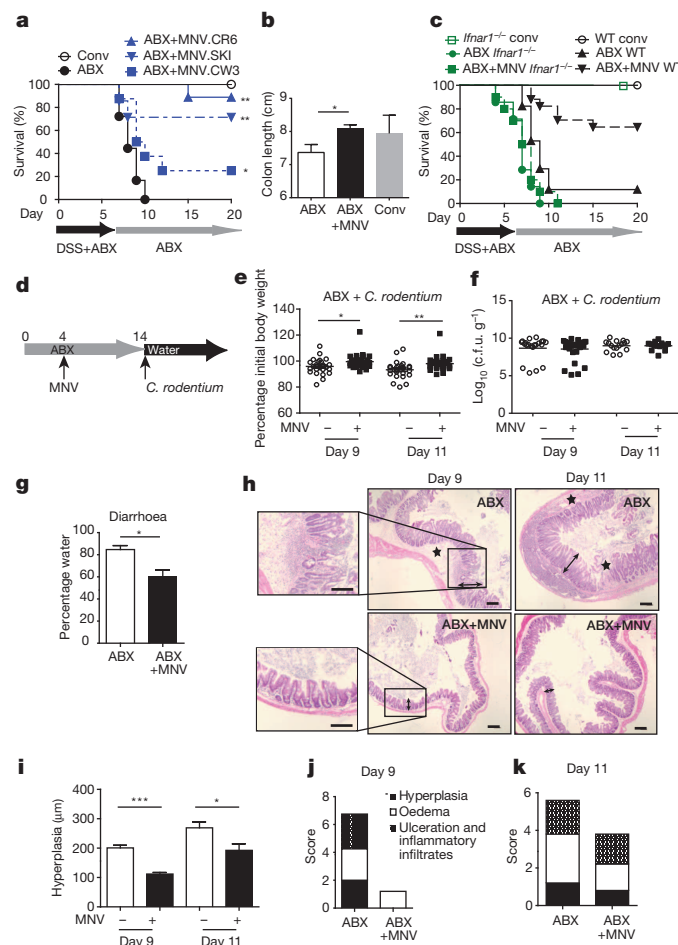


Figure 4 | MNV protects antibiotic-treated mice from intestinal injury and *C. rodentium* superinfection. **a**, Survival after DSS treatment of conventional mice (Conv), mice receiving antibiotics (ABX) only, and mice receiving antibiotics and infected with indicated MNV strains. $n = 12$ mice per group. Statistical significance represents comparisons between antibiotic-treated mice infected with indicated MNV strains versus the antibiotics-only group. **b**, Colon length of conventional, antibiotics-only and antibiotics plus MNV.CR6 mice on day 6 of DSS treatment. $n = 3$ –5 mice per group. **c**, Survival of conventional wild-type (WT) and *Ifnar1*^{-/-} mice, and antibiotic-treated wild-type and *Ifnar1*^{-/-} mice with and without MNV.CR6 infection, treated with DSS. $n = 7$ mice per group. **d**, Schematic of MNV.CR6 and *C. rodentium* co-infection in antibiotic-treated mice. Antibiotic treatment was stopped before *C. rodentium* infection. **e**, **f**, Quantification of weight loss (**e**) and c.f.u. in stool (**f**) in mice receiving treatment as illustrated in **d** on day 9 and 11 after infection with *C. rodentium*. $n = 5$ mice per group. **g**, Diarrhoea from the mice described earlier on day 9 after infection. **h**–**k**, Representative H&E-stained caecal sections (**h**) from the mice described earlier (arrows span hyperplasia and stars denote ulceration and oedema), quantification of hyperplasia (**i**), and histopathology score at day 9 and 11 (see Methods) (**j**, **k**). Scale bar, 150 μ m ($n = 5$ mice per group). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. **a**, **b**, **e**, **g**, **i**, Log-rank Mantel–Cox test (**a**), ANOVA with Holm–Sidak correction (**b**, **e**, **i**), unpaired two-tailed t -test (**g**). Bar graphs show means \pm s.e.m. and bars in **e** and **f** represent mean from at least two independent experiments.

5. Donaldson, E. F., Lindesmith, L. C., Lobue, A. D. & Baric, R. S. Norovirus pathogenesis: mechanisms of persistence and immune evasion in human populations. *Immunol. Rev.* **225**, 190–211 (2008).
6. Popgeorgiev, N., Temmam, S., Raoult, D. & Desnues, C. Describing the silent human virome with an emphasis on giant viruses. *Intervirology* **56**, 395–412 (2013).
7. Ninomiya, M., Takahashi, M., Nishizawa, T., Shimosegawa, T. & Okamoto, H. Development of PCR assays with nested primers specific for differential detection of three human anelloviruses and early acquisition of dual or triple infection during infancy. *J. Clin. Microbiol.* **46**, 507–514 (2008).
8. Ott, C. *et al.* Use of a TT virus ORF1 recombinant protein to detect anti-TT virus antibodies in human sera. *J. Gen. Virol.* **81**, 2949–2958 (2000).
9. Minot, S. *et al.* The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* **21**, 1616–1625 (2011).
10. Modi, S. R., Lee, H. H., Spina, C. S. & Collins, J. J. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499**, 219–222 (2013).
11. Reyes, A. *et al.* Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**, 334–338 (2010).
12. Handley, S. A. *et al.* Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome. *Cell* **151**, 253–266 (2012).
13. Karst, S. M., Wobus, C. E., Lay, M., Davidson, J. & Virgin, H. W. IV STAT1-dependent innate immunity to a Norwalk-like virus. *Science* **299**, 1575–1578 (2003).
14. Wobus, C. E. *et al.* Replication of norovirus in cell culture reveals a tropism for dendritic cells and macrophages. *PLoS Biol.* **2**, e432 (2004).
15. Cadwell, K. *et al.* Virus-plus-susceptibility gene interaction determines Crohn's disease gene *Atg16L1* phenotypes in intestine. *Cell* **141**, 1135–1145 (2010).
16. Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nature Rev. Immunol.* **9**, 313–323 (2009).
17. Kuss, S. K. *et al.* Intestinal microbiota promote enteric virus replication and systemic pathogenesis. *Science* **334**, 249–252 (2011).
18. Kane, M. *et al.* Successful transmission of a retrovirus depends on the commensal microbiota. *Science* **334**, 245–249 (2011).
19. Olszak, T. *et al.* Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* **336**, 489–493 (2012).
20. Thackray, L. B. *et al.* Murine noroviruses comprising a single genogroup exhibit biological diversity despite limited sequence divergence. *J. Virol.* **81**, 10460–10473 (2007).
21. Mortha, A. *et al.* Microbiota-dependent crosstalk between macrophages and ILC3 promotes intestinal homeostasis. *Science* **343**, 1249288 (2014).
22. Wlodarska, M. *et al.* Antibiotic treatment alters the colonic mucus layer and predisposes the host to exacerbated *Citrobacter rodentium*-induced colitis. *Infect. Immun.* **79**, 1536–1545 (2011).
23. Kamada, N. *et al.* Regulated virulence controls the ability of a pathogen to compete with the gut microbiota. *Science* **336**, 1325–1329 (2012).
24. Grohmann, G. *et al.* Outbreak of human calicivirus gastroenteritis in a day-care center in Sydney, Australia. *J. Clin. Microbiol.* **29**, 544–550 (1991).
25. Matson, D. O., Estes, M. K., Tanaka, T., Bartlett, A. V. & Pickering, L. K. Asymptomatic human calicivirus infection in a day care center. *Pediatr. Infect. Dis. J.* **9**, 190–195 (1990).
26. González-Navajas, J. M., Lee, J., David, M. & Raz, E. Immunomodulatory functions of type I interferons. *Nature Rev. Immunol.* **12**, 125–135 (2012).
27. Basic, M. *et al.* Norovirus triggered microbiota-driven mucosal inflammation in interleukin 10-deficient mice. *Inflamm. Bowel Dis.* **20**, 431–443 (2014).
28. Lipkin, W. I. & Firth, C. Viral surveillance and discovery. *Curr Opin Virol* **3**, 199–204 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank S. Koralov and P. Loke for advice on the manuscript, E. Venturini for assistance with deep sequencing, S. Brown and Z. Tang for data analysis, L. Ciriboga for CD3 staining, the flow cytometry and histopathology cores (Cancer Center Support Grant, P30CA016087) for assistance with sample preparation and analyses, M. Alva and D. Littman for assistance with breeding and maintaining GF mice, and H. Moura Silva for sample collection for MNV isolation. This research was supported by National Institutes of Health grant R01 DK093668 (K.C.) and a New York University Whitehead Fellowship (K.C.), Vilcek Fellowship (E.K.) and Erwin Schrödinger Fellowship from the Austrian Science Foundation (E.K.).

Author Contributions E.K. performed all the experiments, Y.D. analysed and scored histological sections, K.C. and E.K. designed the study and wrote the manuscript.

Author Information RNA-seq data have been deposited in the Gene Expression Omnibus under accession number GSE60163. The MNV.SKI capsid sequence has been deposited in the NCBI Reference Sequence database under accession number KM463105. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.C. (ken.cadwell@med.nyu.edu).

METHODS

Mice. GF mice on a C57BL/6J background were bred and housed in flexible film isolators in the New York University (NYU) School of Medicine Gnotobiotics Animal Facility²⁹. Absence of faecal bacteria and fungi was confirmed by aerobic culture in brain heart infusion, sabaraud and nutrient broth (Sigma) and qPCR for bacterial 16S (UniF340 5'-ACTCCTACGGGAGGCAGCACT-3'; UniR514 5'-ATTACCGCGTCTGCTGGC-3') and eukaryotic 18S ribosomal RNA genes (B2F 5'-ACTTCGATGGTAGGATAG-3'; B4R 5'-TGATCGTCTTCGATCCCCTA-3') through sampling of stool from individual cages in each isolator on a monthly basis. Mono-association with MNV as described in Extended Data Fig. 1 was performed within a dedicated isolator by inoculating three male and three female GF mice with 3×10^6 p.f.u. of MNV.CR6 from an endotoxin-free stock prepared as described later. MNV mono-associated GF mice (GF+MNV) were bred to each other (F_0 generation), and successful transmission to offspring was confirmed by performing a plaque assay on stool collected from 8–10-week-old adult progeny (4–5 weeks after weaning) (Extended Data Fig. 1). Progeny (F_1) was used for experimental analyses or further breeding to generate additional mice for analyses (F_2). Absence of bacteria and fungi was confirmed as described earlier. Control GF mice were bred and maintained in a separate isolator free of MNV. Unless stated otherwise, GF and GF+MNV mice remained in isolators until the time of analyses. Conventional C57BL/6J wild-type and *Ifnar1*^{-/-} (B6.129S2-*Ifnar1*^{tm1Agt}/Mmjax) mice were purchased from Jackson Laboratories and bred onsite in a MNV and *Helicobacter*-negative specific-pathogen free (SPF) animal facility. The absence of segmented filamentous bacteria (SFB) was determined by PCR (SFB736F 5'-GACGCTGAGGCATGAGAGCAT-3'; SFB844R 5'-GACGGCACGGATTGTTATTCATCA-3'). Age (6–9 weeks) and gender-matched mice were used in all experiments and assigned randomly to experimental groups. Sample size for animal experiments was chosen based on prior data generated in the laboratory. All animal studies were performed according to protocols approved by the NYU School of Medicine Institutional Animal Care and Use Committee (IACUC).

Antibiotics and DSS treatment. Mice were gavaged with 100 mg streptomycin (Sigma) and the drinking water was immediately replaced with filter-sterilized water containing ampicillin (1 g l^{-1} ; American bioanalytical), vancomycin (0.5 g l^{-1} ; MP biomedical), neomycin (1 g l^{-1} ; Sigma), metronidazole (1 g l^{-1} ; Sigma) and 1% sucrose (Fisher). Antibiotic-containing water was replaced at least once a week. For DSS experiments, mice were given 4% DSS (TdB consultancy) with or without the antibiotics cocktail described earlier in their drinking water for 6 days and then received regular or antibiotic-containing drinking water for the remainder of the experiment.

MNV and bacterial inoculation. MNV.CR6 and MNV.CW3 concentrated stock were prepared as described³⁰. Briefly, supernatant from 293T cells transfected with a plasmid containing the viral genome was applied to RAW264.7 cells to amplify virus production, and virions were concentrated by ultracentrifugation and resuspension in endotoxin-free PBS. Concentration of stock was determined by plaque assay as previously described³¹. For indicated experiments in which adult GF mice were exposed to MNV or bacteria outside of an isolator, 6- to 8-week-old GF mice were transferred to pre-assembled autoclaved cages with autoclaved food and water. Mice were infected orally by pipette with 3×10^6 p.f.u. resuspended in 25 μl PBS or gavaged with 100 μl of a suspension of a fresh faecal pellet from a conventional C57BL/6J mouse homogenized in 1 ml PBS. *B. thetaiotaomicron* VPI-5482 provided by E. Martens was grown in peptone yeast glucose broth for 36 h in an anaerobic chamber (Anaerobe Systems). Mice were gavaged with 100 μl of culture containing 10^6 – 10^7 c.f.u. and colonization was confirmed by plating stool, small intestinal tissue and colonic tissue on bacteroides bile esculin (BBE agar, Anaerobe Systems) plates. *L. johnsonii* was provided by R. Jenq and grown in MRS broth (BD) for 16 h under anaerobic conditions. Mice were inoculated with 100 μl of culture (10^7 c.f.u.) and stool, small intestinal and colonic tissue was plated on MRS agar (BD) for enumeration of *L. johnsonii* infection. *C. rodentium* DBS100 was grown in LB to $\text{OD}_{600 \text{ nm}} = 1$, resuspended in PBS, and mice were gavaged with 100 μl containing 2×10^9 c.f.u. Bacterial inocula were determined by serial dilution plating on MacConkey plates. Stool samples collected on indicated days were homogenized in 1 ml PBS, serially diluted and plated on MacConkey plates for quantification of bacterial burden. Diarrhoea was quantified by subtracting the stool's dry weight from its wet weight and expressing the proportion of water as a percentage of the wet weight. One-hundred micrograms poly(I:C) (high molecular weight, InvivoGen) was injected intraperitoneally daily for 10 days starting on day 4 of antibiotic treatment. In cases where antibiotics were discontinued before bacterial inoculation, stool samples from individual cages were collected at the last day of antibiotic treatment and at the end-point of *C. rodentium*, *B. thetaiotaomicron* or *L. johnsonii* infection, DNA was extracted and qPCR was performed for the 16S genes of the following bacterial groups: *Lactobacilli* (Fwd 5'-AGCAGTAGGGAATCTTCCA-3', Rev 5'-CACCGCTACACA TGGAG-3'); *Enterobacteriaceae* (Uni515F 5'-GTGCCAGCMGCCGCGGTAA-3'; Ent826R 5'-CCTCAAGGGCACAACCTCCAAG-3'); *Fusobacterium* (Fwd 5'-CT AACGCGATAAGTAATC-3', Rev 5'-TGGTAACATACGATAGGG-3'); *Bacteroides*

(Fwd 5'-GAGAGGAAGGTCCCCAC-3', Rev 5'-CGCTACTTGGCTGGTTCA G-3'); and *C. rodentium* (Fwd 5'-AGGCCTTCGGGTGTAAAGT-3', Rev 5'-AT TCCGATTAACGCTTGAC-3').

Microscopy. Small intestinal, colonic and caecal tissue was prepared for staining as previously described¹⁵. H&E and anti-CD3 staining were performed by the NYU Histopathology core and NYU IHC core, respectively. Lysozyme staining of small intestinal sections was performed as described³². At least 50 villi per mouse were measured for villi width at the base of the villi where they meet the crypt or were analysed for number of CD3⁺ cells. Paneth cell granules were counted in at least 30 crypts per mouse. Mean values were calculated for each mouse and used as individual data points. Sections were imaged on a Zeiss Axioplan microscope. All analyses of slides were performed blind and quantified using ImageJ software. The histopathology score for individual mice is the sum of three parameters, including villus hyperplasia and crypt architectural distortion, degree of inflammation and oedema thickness in the lamina propria and submucosa. The scoring criteria are as follows: villus hyperplasia and crypt architectural distortion: 0 = no; 1 = mild; 2 = moderate; 3 = severe; inflammation: 0 = no; 1 = focal; 2 = multifocal; 3 = transmural; oedema thickness: 0 = 0 μm ; 1 = 1–100 μm ; 2 = 101–200 μm ; 3 = 201–300 μm , giving a maximal score of 9. The mean of the score for individual mice was calculated. Blind histopathology scoring was performed by a pathologist (Y.D.).

Flow cytometry. Small intestinal or colonic tissue was flushed with PBS, fat and Peyer's patches were removed and the tissue was incubated in HBSS with 5 mM EDTA and 1 mM dithiothreitol (DTT) at 37 °C for 20 min, followed by HBSS and 5 mM EDTA for 10 min. The tissue was digested using collagenase (Sigma) for 30 min, followed by a Percoll (Fisher) gradient centrifugation using 40% and 80% Percoll. Single-cell suspensions of MLNs were prepared by passing MLNs through 100 μm cell strainers (BD) and subsequent resuspension in PBS. Total cell numbers per ml were determined by counting recovered cells in a haemocytometer. For intracellular cytokine staining, cells were stimulated using the eBioscience cell stimulation cocktail for 4 h at 37 °C. Cells were fixed and permeabilized using the Biolegend fixation and permeabilization buffers. Samples were stained with a fixable live dead stain (Invitrogen) and live gating was performed before any other gating for all samples. The following antibodies (clones) were used for staining: CD4 (GK1.5), CD8 (53-6.7), CD3 (145-2C11), CD11b (M1/70), TCR- β (H57-597), T-bet (4B10), IFN- γ (XMG1.2), IL-17 (TC11-18H10.1), CD19 (6D5), Fc block (TruStainfcx), all from Biolegend; GATA3 (L50-823) from BD bioscience and IL-22 (1H8PWSR), IL-13 (eBio13A), B220 (RA3-6B2) from eBioscience. The National Institutes of Health tetramer facility provided the mCD1d tetramer. FlowJo v.10 was used to analyse flow cytometry data.

RNA-seq. RNA was extracted from 2 cm of snap-frozen small intestinal tissue with TRIzol reagent (Fisher Bioscience) using the manufacturer's protocol. An RNA library was prepared using the Illumina TruSeq RNA sample preparation kit and sequenced with the Illumina HiSeq2000 using the TruSeq RNA v.2 protocol. Illumina CASAVA v.1.8.2 was used to generate FASTQ files containing 29.5–53.5 million qualified reads per sample. Alignment and gene expression count were computed using default settings, which aligns reads to the union of all RefSeq-annotated exons for each gene. Fifty-nine per cent to 78% of reads were aligned to mouse genome version mm9. All genes expressed below one base pair per million (c.p.m.) in the majority of replicates in all treatments were filtered from the data set. The Bioconductor package edgeR was used to find differentially expressed genes with a false discovery rate (FDR) below 5% (ref. 33). The Database of Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics resource was used to assign differentially regulated genes to GO functional categories and determine enrichment scores and significance by a modified Fisher's exact test³⁴. Gene set enrichment analysis (GSEA; Broad institute) was performed in Gene Pattern using the GSEA_Pranked module and a Geneset file created by the Bader laboratory for mouse (http://download.baderlab.org/EM_Genesets/current_release/Mouse/Entrezgene) with RNA-seq gene expression results ranked by fold change of MNV treated versus uninfected and of conventional versus uninfected.

Enzyme-linked immunosorbent assay. Two centimetres of small intestinal tissue was homogenized in 1 ml PBS containing protease inhibitor cocktail or serum and was collected. Samples were diluted 1:100 and levels of IgA and IgG2c (ng ml^{-1}) were determined according to the manufacturer's protocol (Southern Biotech).

Isolation of MNV from vivarium. Stool from *Rag1*^{-/-} mice housed in a MNV-positive room within the Skirball vivarium was tested for the presence of MNV with previously described primers and nested PCR condition³⁵. Stool from the same mice was homogenized in 1 ml of DMEM, and the supernatant was cleared by centrifugation and filtered through a 0.2 μm filter. Fifty microlitres were added to a well of RAW264.7 cells seeded in a 6-well dish. Five days later the cells and supernatant were harvested, freeze-thawed and transferred to a 10 cm dish of RAW264.7 cells for amplification. After another 5 days, virus in supernatant was concentrated by ultracentrifugation. For phylogenetic analysis RNA was isolated using TRIzol and standard protocols and complementary DNA was prepared with ProtoScript M-MuLV First

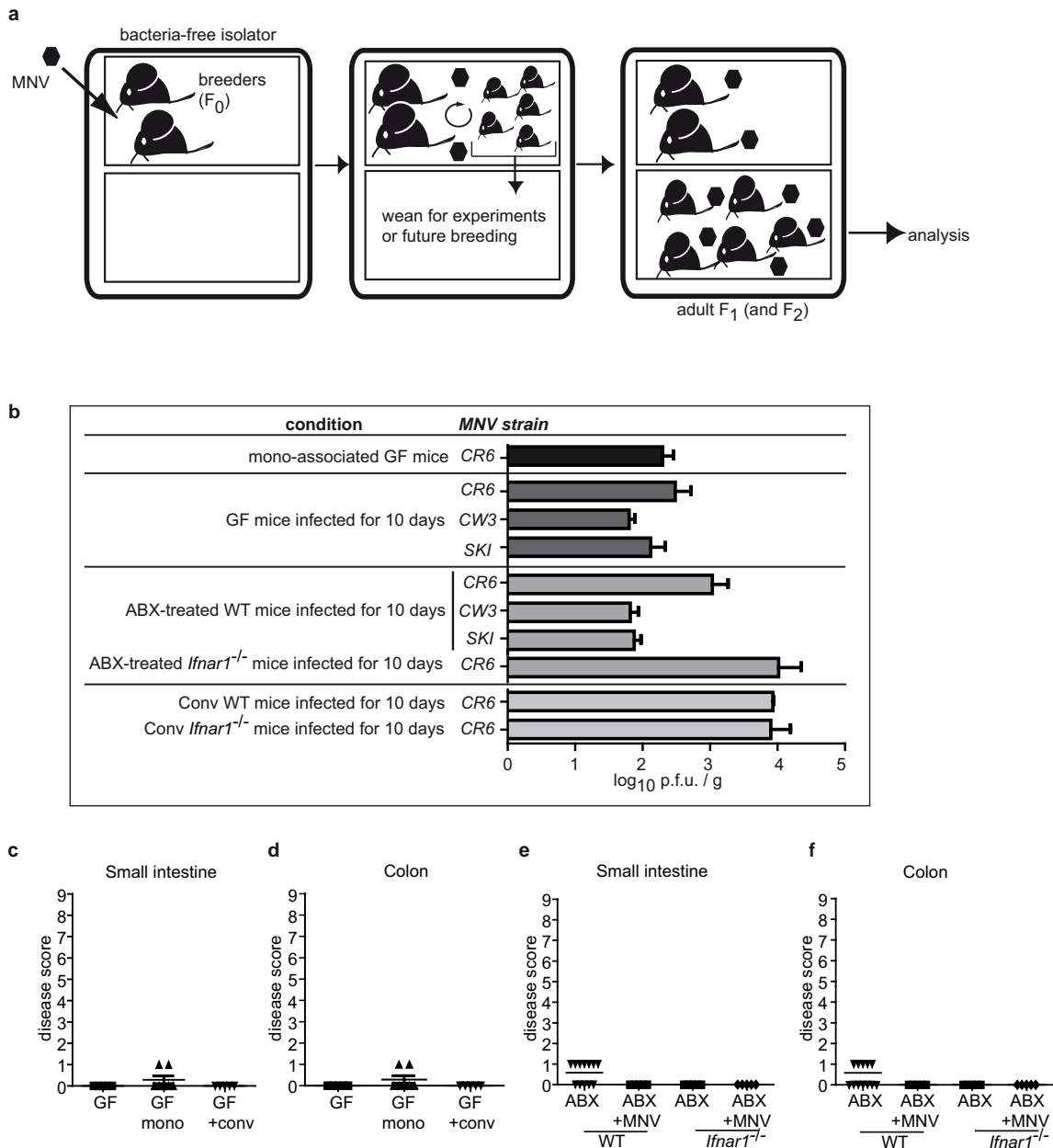
Strand cDNA synthesis kit (New England Biolabs) and random hexamers. The capsid region was amplified and sequenced as previously described³⁶. Capsid sequences from various MNV strains were aligned using Clustal Omega (EMBL-EBI).

qRT-PCR. Total RNA from colonic tissue was isolated using TRIzol and standard protocols, followed by cDNA synthesis using ProtoScript M-MuLV First Strand cDNA synthesis kit (New England Biolabs) and OligodTs. qPCR was performed on a Roche480II Lightcycler using the following primers: *Gbp2*, Fwd 5'-TGCTAAACTTCGGGAACAGG-3', Rev 5'-GAGCTTGGCAGAGAGGTTTG-3'; *Il10*, Fwd 5'-GCCGTCAATTTCTGCCTCAT-3', Rev 5'-GCTTCCCTATGGCCCTCATT-3'; *Ifng*, Fwd 5'-ATGAACGCTACACACTGCATC-3', Rev 5'-CCATCCTTTTGCCAGTTCCCTC-3'; *Gapdh*, Fwd 5'-TGCCCCCATGTTTGTGATG-3', Rev 5'-TGTGGTCATGAGCCCTTCC-3'. Relative expression of the respective genes to *Gapdh* expression was calculated using the $\Delta\Delta C_T$ method and values were expressed as fold change from antibiotic-treated mice. Total RNA from stool was isolated with TRIzol followed by cDNA synthesis with random hexamer primers for quantification of *C. rodentium* virulence factor expression with the following primers: *ler*, Fwd 5'-AATATACCTGATGGTGCTCTTG-3', Rev 5'-TTCTTCCATTCAATAATGCTTCTT-3'; *tir*, Fwd 5'-TACACATTCGGTTATTCAGCAG-3', Rev 5'-GACATCCAACCTTCAGCATA-3'; and *recA*, Fwd 5'-CGCATTCGCTTTACCCTGACC-3', Rev 5'-TCGTCGAAATCTACGGACCGGA-3'.

Statistical analysis. Analyses except for RNA-seq data used Graphpad Prism v.6. An unpaired two-tailed *t*-test was used to evaluate differences between two groups where data was distributed normally with equal variance between conditions. An analysis of variance (ANOVA) with Holm-Sidak multiple comparisons test was used

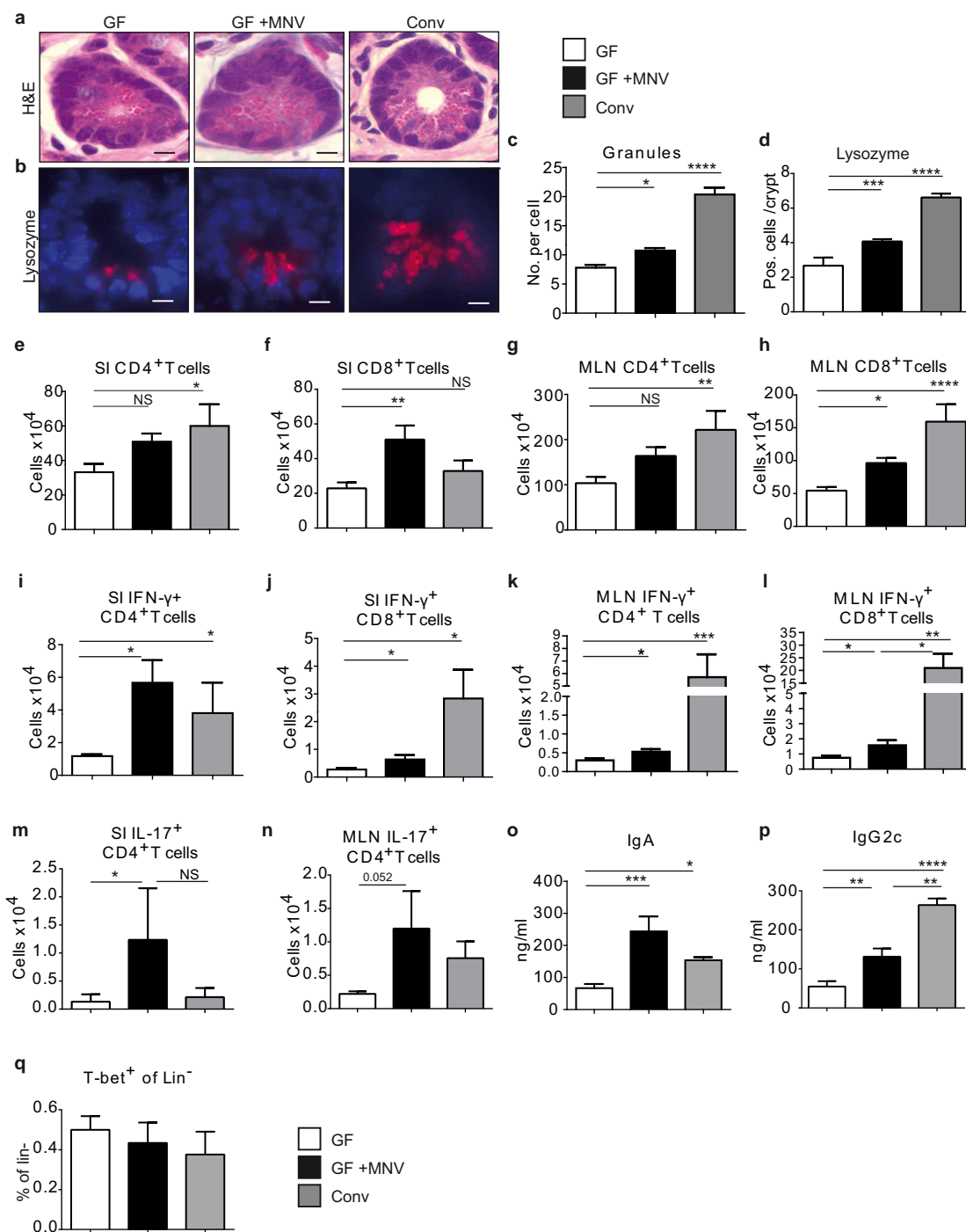
to evaluate experiments involving multiple groups. The log-rank Mantel-Cox test was used for comparison of mortality curves.

29. Marchiando, A. M. *et al.* A deficiency in the autophagy gene *Atg16L1* enhances resistance to enteric bacterial infection. *Cell Host Microbe* **14**, 216–224 (2013).
30. Strong, D. W., Thackray, L. B., Smith, T. J. & Virgin, H. W. Protruding domain of capsid protein is necessary and sufficient to determine murine norovirus replication and pathogenesis *in vivo*. *J. Virol.* **86**, 2950–2958 (2012).
31. Gonzalez-Hernandez, M. B., Bragazzi Cunha, J. & Wobus, C. E. Plaque assay for murine norovirus. *J. Vis. Exp.* e4297 (2012).
32. Cadwell, K. *et al.* A key role for autophagy and the autophagy gene *Atg16l1* in mouse and human intestinal Paneth cells. *Nature* **456**, 259–263 (2008).
33. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
34. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).
35. Tajima, M., Kotani, Y., Kurosawa, T. & Miyasaka, M. Pitfalls in mouse norovirus (MNV) detection in fecal samples using RT-PCR, and construction of new MNV-specific primers. *Exp. Anim.* **62**, 127–135 (2013).
36. Kim, M., Lee, H., Chang, K. O. & Ko, G. Molecular characterization of murine norovirus isolates from South Korea. *Virus Res.* **147**, 1–6 (2010).
37. Izcue, A. *et al.* Interleukin-23 restrains regulatory T cell activity to drive T cell-dependent colitis. *Immunity* **28**, 559–570 (2008).
38. Mumfrey, S. M. *et al.* Murine norovirus 1 infection is associated with histopathological changes in immunocompetent hosts, but clinical disease is prevented by STAT1-dependent interferon responses. *J. Virol.* **81**, 3251–3263 (2007).



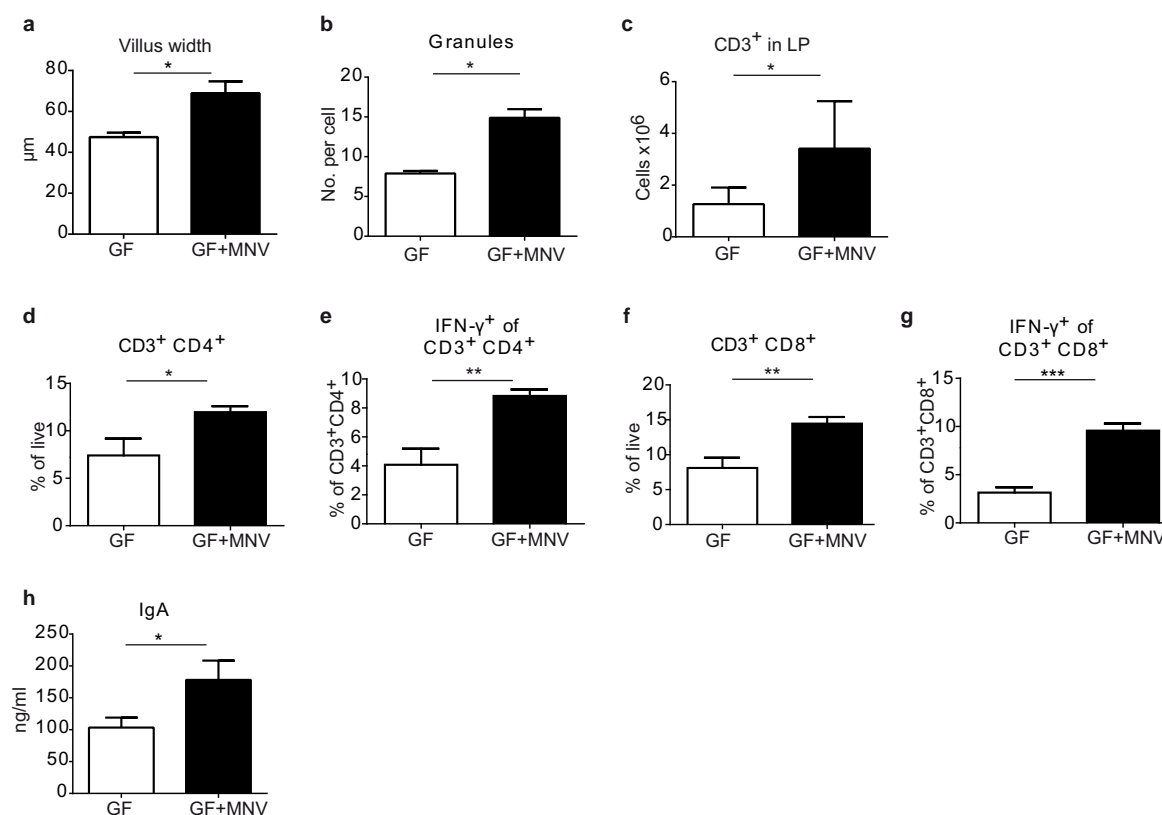
Extended Data Figure 1 | Mono-association of GF mice with MNV does not lead to uncontrolled viral replication or disease. **a**, Schematic of MNV mono-association procedure (see Methods for additional details). GF breeder pairs (F_0) within a gnotobiotic isolator were infected with 3×10^6 p.f.u. of MNV.CR6, which was allowed to transmit naturally to offspring (F_1). Weaned offspring were maintained in isolators until adulthood, and then either used for analysis or as breeders to generate additional experimental animals (F_2). **b**, Successful transmission to offspring and the persistent presence of the virus in mono-associated GF mice (GF+MNV) was confirmed by performing a plaque assay using stool harvested from 8-week-old offspring (~ 1 month after weaning). The amount of virus in stool from GF mice, antibiotic (ABX)-treated wild-type (WT) and *Ifnar1*^{-/-} mice, and conventional (Conv) mice infected with 3×10^6 p.f.u. of the indicated strains of MNV for 10 days are also

shown, $n = 5$ mice per group. **c–f**, Mice receiving the indicated treatments did not display marked histopathology in the small intestine (**c**, **e**) and colon (**d**, **f**) based on blind quantification of H&E-stained sections using a previously described scoring system³⁷. Mice receiving a pathology score of 1 displayed mild blunting of villi (additional details in Methods). No histopathology was detected in spleens, and no other signs of disease were noted. Note that a previous publication in which mice were reported to display pathologies after MNV infection used a different strain of MNV, an early time point (24 h after infection), and mice on a different background (129/Sv)³⁸. The lack of pathology in C57BL/6 mice persistently infected with MNV.CR6 is consistent with our previous publication¹⁵. $n = 5–7$ mice per group. All graphs show means \pm s.e.m.



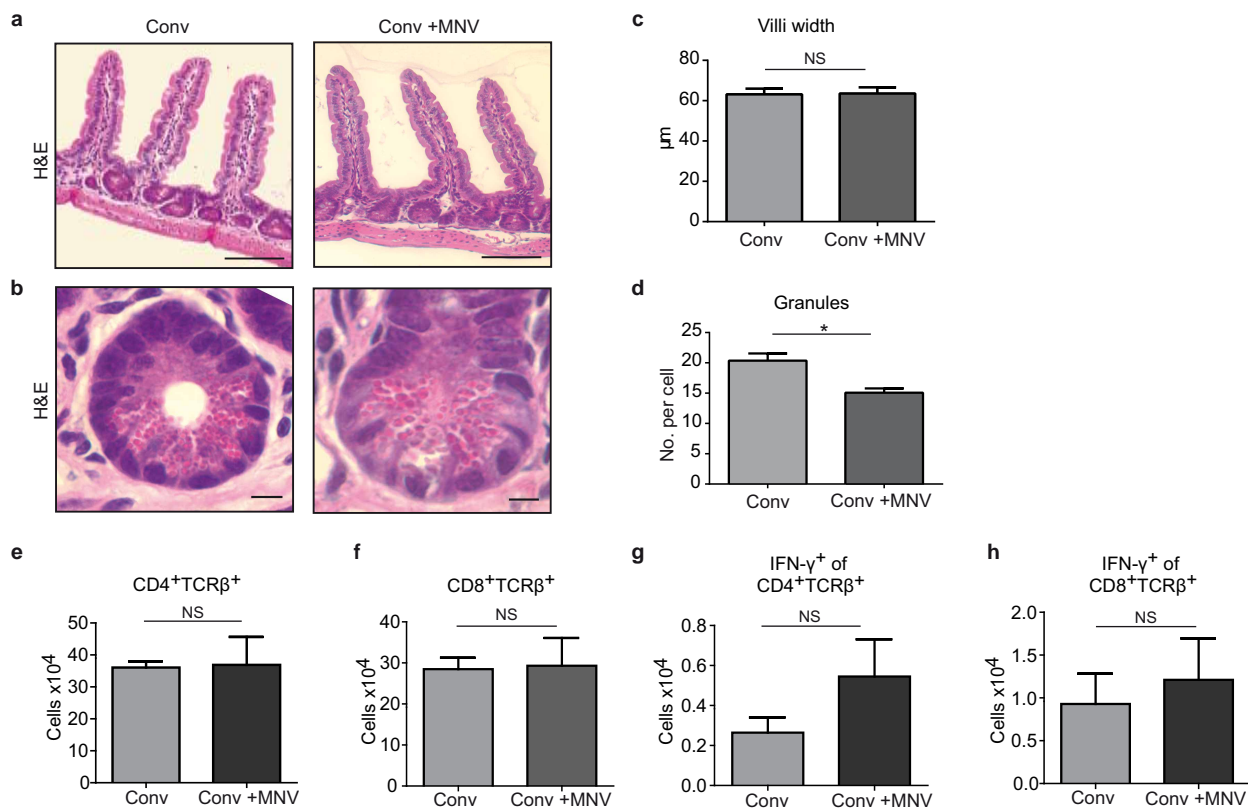
Extended Data Figure 2 | MNV improves several deficiencies related to intestinal immunity in GF mice. **a, b**, Representative images of crypts from small intestinal tissue sections stained with H&E (**a**) and an anti-lysozyme antibody (**b**) harvested from GF, GF+MNV (mono-association) and conventional (Conv) mice. Scale bar, 1 μ m. **c, d**, Quantification of the above images shows an increase in granules per Paneth cell (**c**) and lysozyme-positive cells per crypt (**d**) in GF+MNV mice, indicating that the presence of MNV partially reverses Paneth cell abnormalities due to the absence of bacteria. $n = 5$ mice per group. **e–h**, MNV mono-association of GF mice increases the number of CD4⁺ (**e, g**) and CD8⁺ (**f, h**) T cells (TCR- β ⁺) in small intestinal (SI) lamina propria cells and MLNs. **i–n**, Flow cytometry analysis indicates that MNV

mono-association of GF mice also increases the number of IFN- γ -expressing CD4⁺ and CD8⁺ T cells in small intestine lamina propria (**i, k**) and MLNs (**j, l**). IL-17 expression by CD4⁺ T cells is also influenced by the presence of MNVs (**m, n**). $n = 10$. **o, p**, GF+MNV mice display increased IgA levels in small intestine tissue (**o**) and IgG2c levels in serum (**p**). $n = 5$ mice per group. **q**, Percentage of T-bet⁺ cells in the small intestine lamina propria after gating on live and Lin⁻ cells remain unchanged by MNV infection of GF mice. $n = 10$ mice per group. NS, not significant. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. ANOVA with Holm–Sidak correction. All graphs show means \pm s.e.m. from at least two independent experiments.



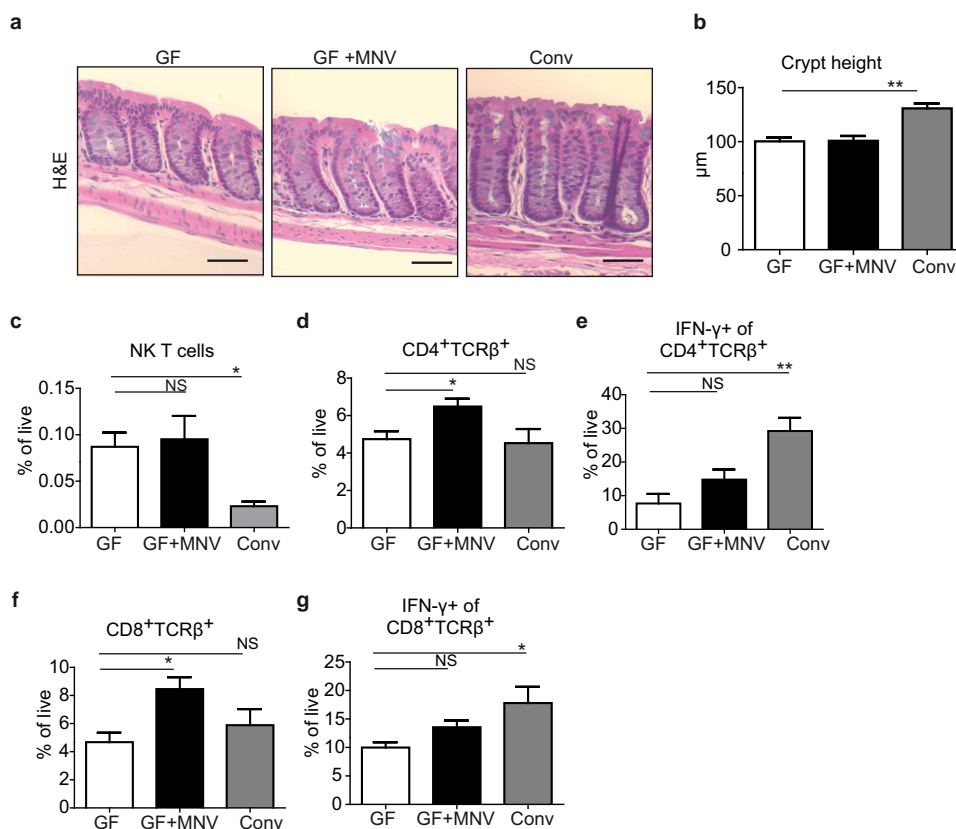
Extended Data Figure 3 | MNV infection of adult GF mice has similar effects to mono-association of GF mice from birth. **a–f**, Six- to eight-week old adult GF mice were infected with MNV.CR6 for 10 days and examined for reversal of intestinal abnormalities. **a**, **b**, Quantification of villus width (**a**) and granules per Paneth cell (**b**) in H&E-stained small intestinal sections. $n = 5$ mice per group. **c**, Quantification of the number of CD3⁺ T cells in the small

intestine lamina propria (LP) by flow cytometric analysis. $n = 6$ mice per group. **d**, **e**, Quantification of IFN- γ -producing CD3⁺CD4⁺ (**d**) and CD3⁺CD8⁺ (**e**) T cells by flow cytometry. **f**, Quantification of small intestinal IgA by enzyme-linked immunosorbent assay (ELISA). $n = 6$ mice per group. * $P < 0.05$. Unpaired two-tailed t -test. All graphs show means \pm s.e.m. from at least two independent experiments.



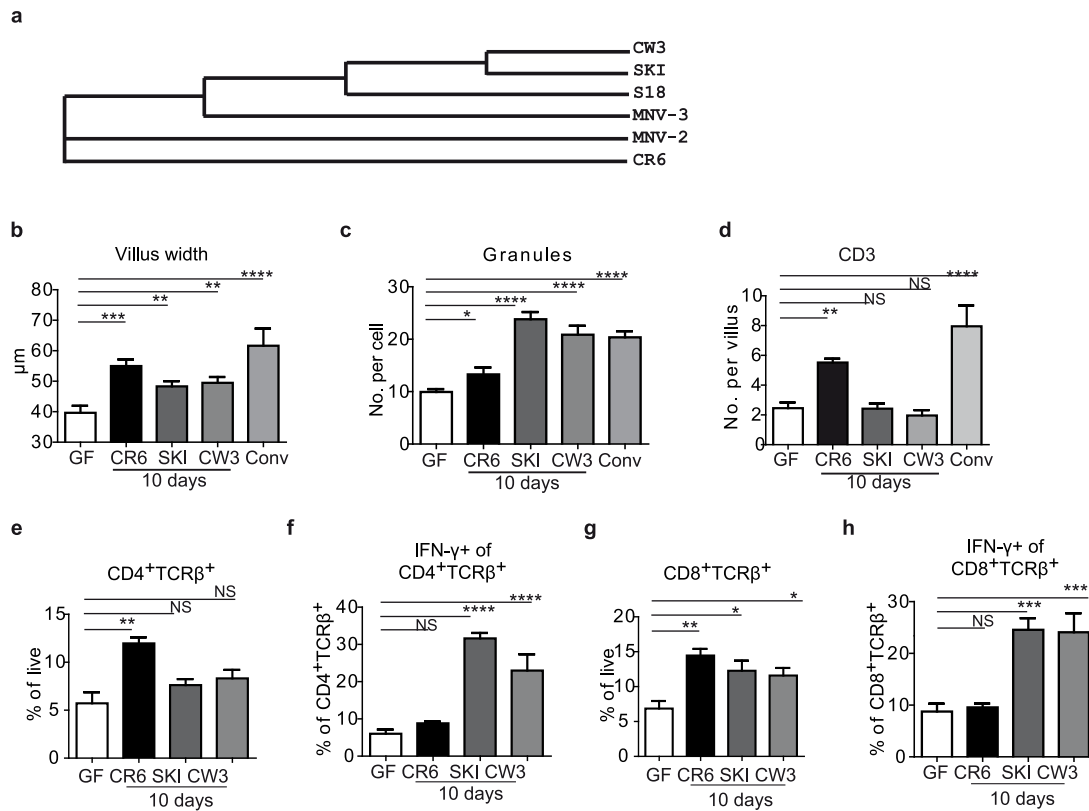
Extended Data Figure 4 | The effect of MNV is specific to mice depleted of bacteria. **a–h**, Conventional mice were infected with 3×10^6 p.f.u. of MNV.CR6 to determine the effect of MNV in the presence of commensal bacteria. **a, b**, Representative images of H&E-stained small intestinal sections of conventional (Conv) and Conv mice infected with MNV (Conv+MNV) mice showing no aberrant changes after MNV infection. Scale bar, 100 μ m (**a**);

1 μ m (**b**). **c, d**, Villus width (**c**) and Paneth cell granules (**d**) were quantified from at least 50 villi and 30 crypts of 2–5 mice per group. **e–h**, Cell numbers of CD4⁺TCR- β ⁺ (**e**), CD8⁺TCR- β ⁺ (**f**), IFN- γ producing CD4⁺ (**g**) and IFN- γ producing CD8⁺ T cells in small intestine lamina propria (**h**). $n = 6$ mice per group. NS, not significant. * $P < 0.05$. Unpaired two-tailed t -test. All graphs show means \pm s.e.m. from at least two independent experiments.



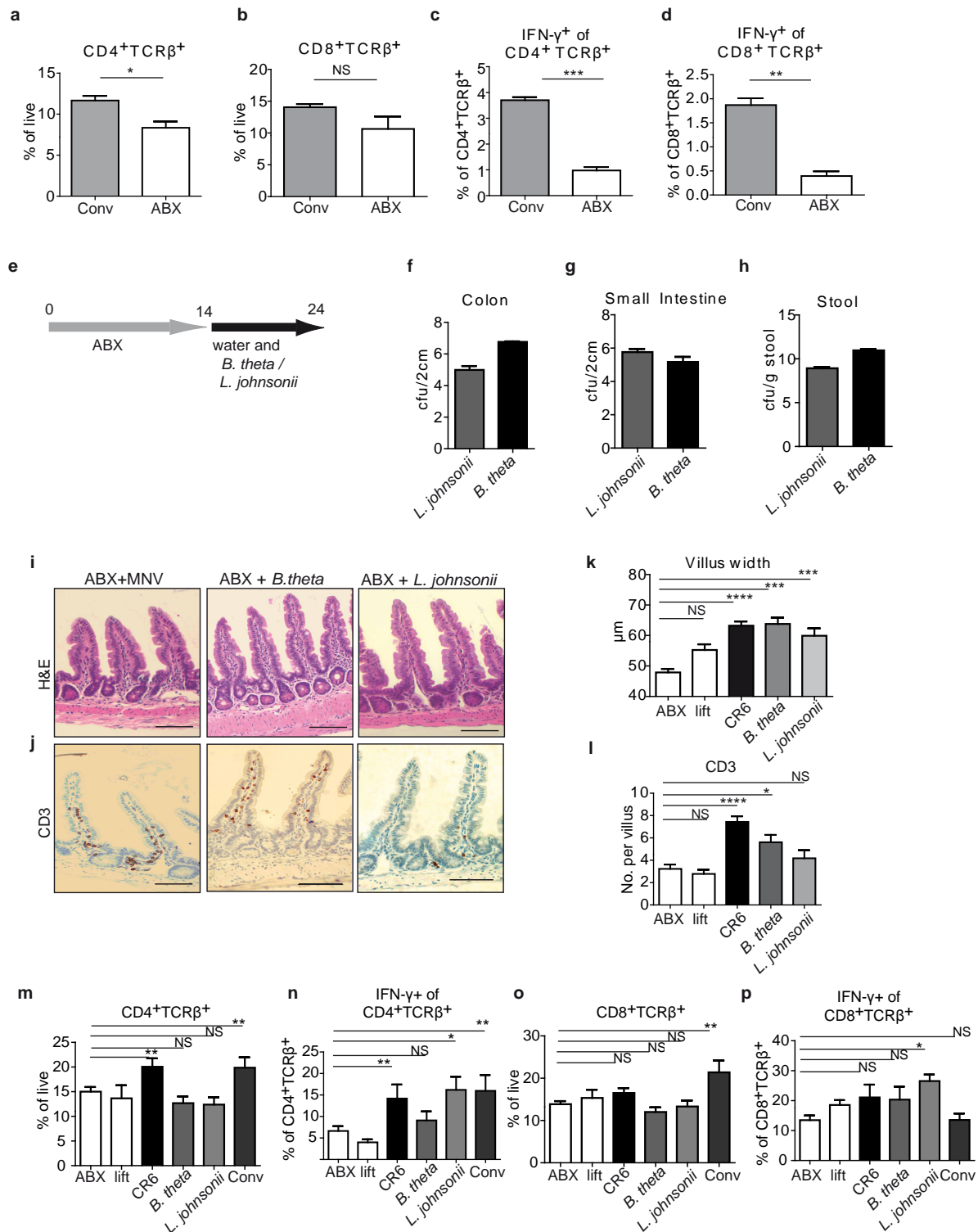
Extended Data Figure 5 | MNV mono-association of GF mice increases colonic lymphocyte populations. **a**, Representative images of H&E-stained colonic small intestinal sections of GF, GF+MNV (mono-association with MNV.CR6) and conventional (Conv) mice. Scale bar, 100 μm . **b**, In these mice the crypt height was measured, showing a significant difference between GF and conventional mice. **c**, Percentages of NK T cells (CD1d^+ , $\text{TCR}\beta^+$) in colonic lamina propria of GF, GF+MNV and conventional mice.

d–g, Percentages of $\text{CD4}^+\text{TCR}\beta^+$ (**d**) and $\text{CD8}^+\text{TCR}\beta^+$ cells (**f**), and percentages of IFN- γ -producing CD4^+ (**e**) and CD8^+ T cells (**g**) in the colonic lamina propria of GF, GF+MNV and conventional mice. $n = 5$ mice per group. NS, not significant. $*P < 0.05$, $**P < 0.01$. ANOVA with Holm–Sidak correction. All graphs show means \pm s.e.m. from at least two independent experiments.



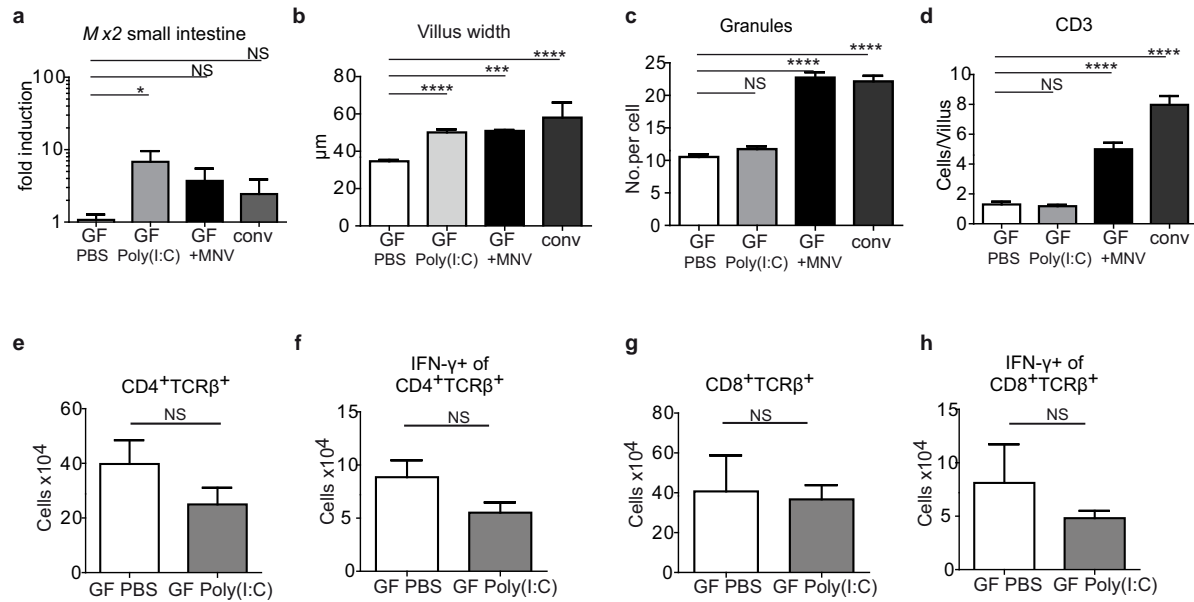
Extended Data Figure 6 | The effect of MNV on the small intestine of GF mice is not strain specific. **a**, Phylogenetic tree of the capsid sequences of the indicated MNV strains. **b–d**, Quantification of the villus width (**b**), granules (**c**) and CD3⁺ cells (**d**) in small intestinal sections prepared from conventional (Conv) mice, GF mice, and GF mice infected with the indicated strains of MNV for 10 days. **e–h**, Percentages of CD4⁺TCR-β⁺ (**e**) and CD8⁺TCR-β⁺ cells (**g**),

and percentages of IFN-γ-producing CD4⁺ (**f**) and CD8⁺ T cells (**h**) in the small intestine lamina propria of the indicated mice. $n = 6$ mice per group. NS, not significant. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. ANOVA with Holm–Sidak correction. All graphs show means \pm s.e.m. from at least two independent experiments.



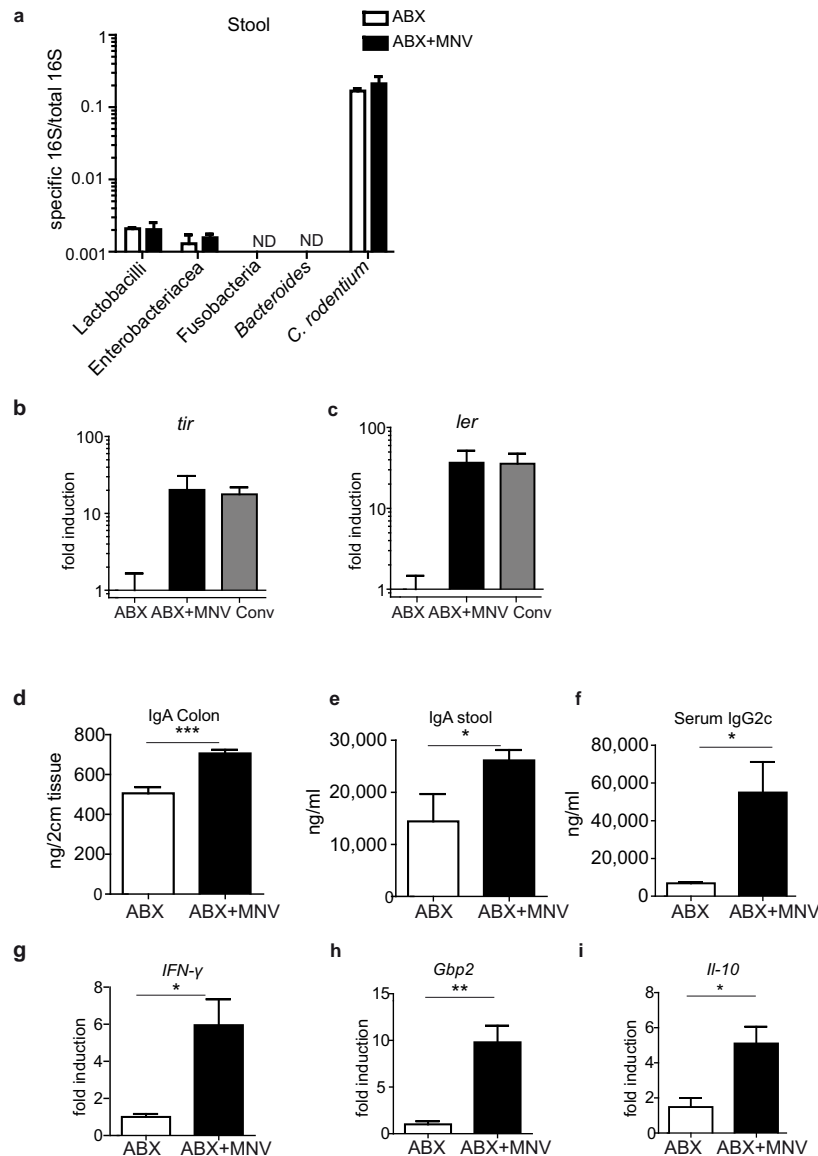
Extended Data Figure 7 | Antibiotics treatment induces intestinal abnormalities that can be reversed by *B. thetaiotaomicron*, *L. johnsonii* or MNV.CR6. **a–g**, Percentages of $CD4^{+}TCR\beta^{+}$ (**a**) and $CD8^{+}TCR\beta^{+}$ cells (**b**), and percentages of IFN- γ -producing $CD4^{+}$ (**c**) and $CD8^{+}$ T cells (**d**) in small intestine lamina propria in conventional (Conv) mice with and without antibiotic (ABX) treatment. *n* = 5–10 mice per group. **e**, Schematic of antibiotic treatment for introducing bacteria. After 14 days, antibiotic-containing water was replaced by regular water. Mice were then inoculated with *B. thetaiotaomicron* (*B. theta*), *L. johnsonii*, or left untreated for 10 days before analyses. **f–h**, Bacterial loads of colon (**f**), small intestine (**g**) and stool (**h**) of antibiotic-treated mice inoculated with *B. thetaiotaomicron* or *L. johnsonii* for 10 days. **i, j**, Small intestinal sections stained with H&E (**i**)

or anti-CD3 antibody (**j**) indicating that inoculation with MNV.CR6, *B. thetaiotaomicron* or *L. johnsonii* have similar effects on intestinal morphology. **k–p**, Mice that received antibiotics during the whole course of the experiment with or without MNV (ABX and ABX+MNV) were compared with mice treated as in **e** using the previously described measurements: quantification of villus width (**k**), CD3⁺ cells per villi (**l**), and percentages of $CD4^{+}$ (**m**), $CD8^{+}$ (**o**), IFN- γ^{+} $CD4^{+}$ (**n**) and IFN- γ^{+} $CD8^{+}$ (**p**) T cells in the small intestine lamina propria. *n* = 8 mice per group. NS, not significant. **P* < 0.05, ***P* < 0.01, ****P* < 0.001. **a–d, k–p**, Unpaired two-tailed *t*-test (**a–d**), ANOVA with Holm–Sidak correction (**k–p**). All graphs show means \pm s.e.m. from at least two independent experiments.



Extended Data Figure 8 | Type I IFN induction by poly(I:C) changes small intestinal architecture without affecting the T-cell compartment. **a**, qRT-PCR quantification of the IFN-I inducible gene *Mx2* in small intestinal tissue of untreated GF mice, GF mice injected for 10 days with poly(I:C), MNV.CR6 mono-associated GF mice (GF+MNV), and conventional (Conv) mice indicate that the poly(I:C) injection procedure induces an IFN-I response. Values represent fold induction of *Mx2* compared with untreated GF mice after normalizing to *Gapdh*. **b–d**, Quantification of the villus width (**b**),

granules (**c**) and CD3⁺ cells in the small intestine of the same mice as in **a** of H&E-stained small intestinal sections. **e–h**, Percentages of CD4⁺TCR-β⁺ (**e**), CD8⁺TCR-β⁺ cells (**g**) and IFN-γ-producing CD4⁺ (**f**) and CD8⁺ (**h**) T cells from the small intestine lamina propria. *n* = 6 mice per group. NS, not significant. ****P* < 0.001, *****P* < 0.0001. **a–h**, ANOVA with Holm–Sidak correction (**a–d**), unpaired two-tailed *t*-test (**e–h**). All graphs show means ± s.e.m. from at least two independent experiments.



Extended Data Figure 9 | MNV infection alters the host response to a super infection with *C. rodentium*. **a**, Taxon-specific 16S qRT-PCR for *Lactobacilli*, *Enterobacteriaceae*, *Fusobacteria*, *Bacteroides* and *C. rodentium* normalized to total 16S gene expression in stool of mice infected for 9 days with *C. rodentium* after antibiotic (ABX) pre-treatment with or without MNV. CR6 infection, showing similar colonization of mice throughout the groups. **b**, **c**, Fold induction of *C. rodentium* virulence factors *tir* (**b**) and *ler* (**c**) compared with the antibiotic group after normalization to *recA* in stool of indicated mice on day 9 after *C. rodentium* infection. **d**–**f**, MNV. CR6 infection

of antibiotic-treated mice before *C. rodentium* infection increases IgA levels in colonic tissue (**d**) and stool (**e**), and IgG2c levels in serum (**f**) at day 9 after *C. rodentium* infection. **g**–**i**, At day 9 after *C. rodentium* infection, antibiotics plus MNV mice display elevated expression of *IFN-γ* (**g**), *Gbp2* (**h**) and *IL-10* (**i**) in colonic tissue compared with antibiotics-only mice. $n = 5$ mice per group. ND, not detectable. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Unpaired two-tailed *t*-test. All graphs show means \pm s.e.m. from at least two independent experiments.

Extended Data Table 1 | Genes displaying increased expression after MNV infection or conventionalization of GF mice

a >1.4 fold upregulated in GF+MNV compared to GF

Term	pValue	Genes
immune response	7.15E-10	CSF2, MBL2, IFIH1, TBK1, TNFSF15, CXCL9, OAS3, RSAD2, IL15, SP110, OAS2, TLR7, CCL4, CXCL10, B2M, KLHL6, CCL20, LOC547349, PGLYRP2, OASL2, OASL1, BCL6, MX1, MX2, LTB, DHX58, GBP8, ICOSL, GBP6, IGJ, LAT, IL18BP, LAX1, CXCL13, SERPINA3G, IRF7, VEGFA, H2-OB, OAS1B, CD79B, OAS1A, H2-T24, OAS1G, OAS1E
positive regulation of immune system processes	1.18E-05	MBL2, ITGAL, ICOSL, IL27RA, IKZF1, TBK1, CD247, PTPN22, IL15, TNFRSF4, B2M, CDKN1A, CORO1A, CD37, CD19, KLHL6, LAX1, LCK, BCL6, CD79B, SASH3
sodium ion transport	1.20E-05	SLC12A6, ATP1B1, SLC5A1, ATP1A1, SLC10A5, SLC10A2, TRPM2, SLC23A1, CATSPER4, SLC4A10, SLC6A8, SLC5A4A, SLC5A8, SLC13A1, SLC13A2, SLC5A12
defense response	6.90E-04	MBL2, IFIH1, IL27RA, TBK1, CRP, CXCL9, RSAD2, SP110, TLR7, CCL4, TNFRSF4, LEAP2, CXCL10, B2M, SERPINA1B, CCL20, KLKB1, PGLYRP2, CLEC2H, MX1, MX2, DHX58, B4GALT1, BMP2, LIPA, MAP2K3, LAT, HIF1A, CXCL13
response to virus	7.39E-04	PLSCR1, IFIH1, ISG15, IRF7, OAS1B, RSAD2, OAS1A, MX1, MX2, TLR7
antigen receptor mediated signaling pathway	9.40 E-04	KLHL6, CD19, LAX1, CD247, LCK, PTPN22, CD79B
immune system development	1.48E-03	CSF2, ICOSL, EPAS1, IKZF1, SLC37A4, PTPN22, ZBTB16, SOX6, IL15, PLSCR1, DOCK2, HIF1A, CXCL13, MFSD7B, LCK, VEGFA, BCL6, JAK2, HEP
hemopoietic or lymphoid organ development	1.99E-03	CSF2, EPAS1, IKZF1, SLC37A4, PTPN22, ZBTB16, IL15, SOX6, PLSCR1, DOCK2, HIF1A, CXCL13, MFSD7B, LCK, VEGFA, BCL6, JAK2, HEPH, LTB, RHOF

b >1.4 fold upregulated in GF+conv compared to GF

Term	pValue	Genes
carboxylic acid transport	2.12E-06	SLC27A1, SLC36A1, SLC7A15, PPARD, SLC38A8, XK, SLC7A10, CACNB4, SLC6A18, SLC10A2, SLC19A1, SLC7A14, SLC6A9, SLC26A6, SLC6A8, SLC25A22, SLC27A2, SLC43A1, SLC46A1
amino acid transport	5.06E-05	SLC36A1, SLC7A15, SLC38A8, SLC7A10, XK, CACNB4, SLC19A1, SLC6A18, SLC7A14, SLC6A9, SLC6A8, SLC25A22, SLC43A1, SLC46A1
glucose transport	5.33E-05	PRKAG3, SLC2A8, G6PC, SLC2A3, SLC2A2, SLC5A1, EDN1, SLC37A4, KLF15
response to hypoxia	7.70E-05	ATP1B1, EPAS1, IL18, EDN1, BNIP3, TRF, CITED2, SLC2A8, VEGFA, P2RX2, HIF3A, NOS2, LCT
enzyme linked receptor protein signaling pathway	7.15E-04	GFGR1, WFIKK2, EID2, NOG, ERBB3, SLC2A8, EIF4EBP1, FRS3, PDGFC, SH2B2, SKIL, INSR, FIGF, NR1H3, TXNIP, RET, IRS2, SMAD9, SMAD7, SMAD4, TAB1, EPHA1, NTRK3, CSRN1P, VEGFA, NTRK2, TGFB3, HPGD
vitamin A metabolic process	1.61E-03	RDH9, RBP4, PPARD, LRAT, DHRS9, RDH16, BCO2
acute inflammatory response	2.53E-03	REG3B, ORM1, C3, CFB, EPHX2, VNN1, CFI, REG3G, CFD, TRF, CD163, C8G
blood vessel morphogenesis	2.62E-03	GFGR1, EPAS1, EGFL7, SMAD7, IL18, EDN1, CDH2, CITED2, APOB, HEY1, HAND2, NTRK2, NOTCH4, VEGFA, PLCD3, RHOB, SOX18, NOS3, NOS2, SOX17, FIGF

c >1.4 fold upregulated common genes

Term	pValue	Genes
glucose transport	2.14E-04	PRKAG3, SLC2A2, SLC5A1, SLC37A4, KLF15
erythrocyte differentiation	0.00172446	EPAS1, VEGFA, BCL6, HEPH, SOX6
steroid metabolic process	0.00176271	RDH9, HSD3B3, APOA1, OSBPL3, HMGCS2, OSBPL1A, SLC37A4, SULT2B1
hemopoiesis	0.00573704	EPAS1, SLC37A4, VEGFA, PTPN22, BCL6, HEPH, IL15, SOX6, ZBTB16
homeostasis of number of cells	0.00742149	EPAS1, SLC37A4, VEGFA, BCL6, HEPH, SOX6
immune system development	0.01097433	EPAS1, SLC37A4, VEGFA, PTPN22, BCL6, HEPH, IL15, SOX6, ZBTB16
oxidation reduction	3.14E-04	ACOX2, BCMO1, HSD3B3, PTGR1, CYP3A13, CYP2C44, CYP3A11, CYP4V3, MOSC1, KMO, BBOX1, RDH9, AKR1B7, CYP27A1, HEPH, RDH16, CYP2C38, MOSC1, KMO, BBOX1, RDH9, AKR1B7, CYP27A1, HEPH, RDH16, CYP2C38, HPGD, CYP3A44
myeloid cell differentiation	0.0198445	EPAS1, VEGFA, BCL6, HEPH, SOX6

d

GF+MNV vs GF	NES	GF+conv vs GF	NES
TCR	5.162	membrane trafficking	4.713
EGFR1	5.079	endocytosis	3.828
cytokine signaling in immune system	4.961	signaling by NGF	3.612
class I MHC mediated antigen processing and presentation	4.642	EGFR1	3.525
BCR	3.897	transport to the Golgi and subsequent modification	3.067
interferon gamma signaling	3.883	platelet homeostasis	3.050
growth hormone receptor signaling	3.730	PPAR signaling pathway	3.045
IL12-mediated signaling events	3.675	growth hormone receptor signaling	3.023
ER-phagosome pathway	3.622	cell-cell communication	2.987
chemokine signaling pathway	3.619	hemostasis	2.930
costimulation by the CD28 family	3.562	trans-golgi network vesicle budding	2.923
signaling by NGF	3.528	integration of energy metabolism	2.385
T cell receptor signaling pathway	3.525	MAPK signaling pathway	2.285
CXCR4- mediated signaling events	3.350	IL6	2.202

Genes upregulated more than 1.4 fold in the RNA-seq experiment were subjected to DAVID pathway analysis (Fig. 3), which assigned genes to GO categories. **a–c**, Individual genes are shown with their GO category for GF + MNV (**a**) and for conventionalized GF (GF + conv) mice (**b**), and genes that overlap the GF + MNV and GF + conv conditions are shown (**c**). **d**, The same data set was analysed using gene set enrichment analysis. NES, normalized enrichment score.

The ESCRT machinery regulates the secretion and long-range activity of Hedgehog

Tamás Matusek^{1,2,3*}, Franz Wendler^{1,2,3*}, Sophie Polès^{1,2,3}, Sandrine Pizette^{1,2,3}, Gisela D'Angelo^{1,2,3}, Maximilian Fürthauer^{1,2,3} & Pascal P. Thérond^{1,2,3}

The conserved family of Hedgehog (Hh) proteins acts as short- and long-range secreted morphogens, controlling tissue patterning and differentiation during embryonic development¹. Mature Hh carries hydrophobic palmitic acid and cholesterol modifications essential for its extracellular spreading². Various extracellular transportation mechanisms for Hh have been suggested, but the pathways actually used for Hh secretion and transport *in vivo* remain unclear. Here we show that Hh secretion in *Drosophila* wing imaginal discs is dependent on the endosomal sorting complex required for transport (ESCRT)³. *In vivo* the reduction of ESCRT activity in cells producing Hh leads to a retention of Hh at the external cell surface. Furthermore, we show that ESCRT activity in Hh-producing cells is required for long-range signalling. We also provide evidence that pools of Hh and ESCRT proteins are secreted together into the extracellular space *in vivo* and can subsequently be detected together at the surface of receiving cells. These findings uncover a new function for ESCRT proteins in controlling morphogen activity and reveal a new mechanism for the transport of secreted Hh across the tissue by extracellular vesicles, which is necessary for long-range target induction.

Diffusible signals, such as secreted morphogens, are key players in inter-cellular communication during development. The Hh morphogen has a critical role in patterning the anterior–posterior axis of the developing epithelial wing imaginal discs (WIDs) in *Drosophila*, by specifying the presumptive intervein domain between vein 3 and vein 4 (v3–v4) through both short- and long-range activity¹. Mechanisms proposed to control spreading of Hh include the formation of multimers, loading on filopodia, and packaging into lipoprotein particles^{4–7}. However, the possibility of Hh transport via membranous vesicles released by cells (exovesicles) has never been functionally explored. Exovesicles may be generated by plasma membrane budding or by fusion of multivesicular bodies (MVBs) with the plasma membrane, leading to the release of MVB intraluminal vesicles into the extracellular space^{8–10}. Both plasma membrane budding and the formation of MVB intraluminal vesicles require the severing of thin membrane stalks by a highly conserved specialized machinery, the ESCRT complex³.

In WIDs, Hh is produced and secreted by the epithelial cells of the posterior compartment. It induces a concentration-dependent graded response in anterior cells, through activation of short-range target genes, such as the transcription factor *engrailed* (*en*) and the Hh receptor *patched* (*ptc*), and long-range target genes, such as *decapentaplegic* (*dpp*), the *Drosophila* homologue of TGF- β ¹¹ and *iroquois* (*iro*)¹². The strong cell death and tumorigenic phenotype induced by the lack of ESCRT function in WIDs^{13–16} made it impossible for us to analyse its role in Hh secretion with classical loss-of-function mutants. We therefore used a temporally controllable RNA interference (RNAi) method restricted to Hh-producing cells and developed a ‘tester line’ (Methods) in which increasing Hh production in the posterior cells of the WID induced the ectopic production of *dpp* in the most anterior cells of the WID, leading to severe outgrowth of the anterior compartment (Fig. 1a and Extended Data Fig. 1a–d). Moreover, we also observed an enlargement of the endogenous

dpp expression domain at the anterior–posterior border (10–12 cells wide instead of 6–8 cells wide in the wild type; Fig. 1b, c), leading to an increase in the size of the v3–v4 intervein space of the adult wings (Extended

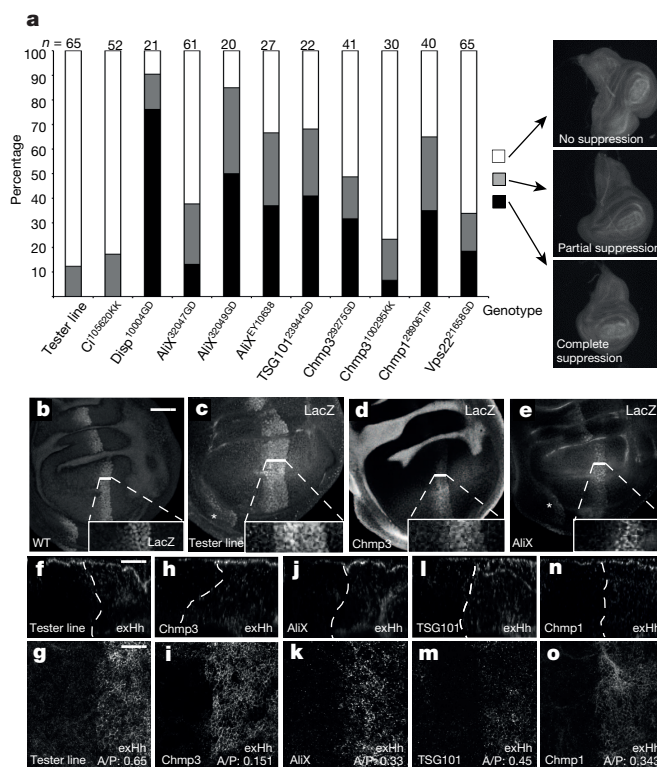


Figure 1 | ESCRT proteins are required for the long-range activity of Hh.

a, Percentage of outgrowth, moderately and fully rescued WIDs with respect to their genotypes. The number of dissected discs (*n*) is shown at the top of the columns. Representative photographs of each category are shown on the right. Negative control: RNAi against *Cubitus interruptus* (Ci), which is not expressed in the posterior compartment. Positive control: RNAi against *Disp*. Superscripts refer to the transformant identity of the RNAi line. **b–e**, *dpp-lacZ* expression in wild-type (WT), tester line and tester line plus ESCRT RNAi discs. A magnification of the *dpp-lacZ* stripe is shown for each genotype in the bottom-right corner. Note that not only the central *dpp* stripe but also the ectopic *dpp* expression is suppressed in tester line plus ESCRT RNAi discs (white asterisks on c and e). Owing to the low level of Dpp staining, signal intensity was increased in d. Scale bar, 50 μ m. **f–o**, Representative X–Z (f, h, j, l, n) and X–Y–Z (g, i, k, m, o) sections showing extracellular Hh staining in control discs and discs with tester line plus ESCRT RNAi. On X–Z sections (scale bar, 20 μ m), apical is towards the top. The anterior–posterior border is labelled with a white dashed line. On X–Y–Z sections (scale bar, 15 μ m), the anterior–posterior (A/P) ratio of Hh staining intensity is indicated in the bottom-right corner for all genotypes. In all images anterior is to the left and in b–e dorsal is to the top. Information about optical sections shown in all figures can be found in Supplementary Table 3. ex, extracellular.

¹Université de Nice Sophia Antipolis, iBV, UMR 7277, 06100 Nice, France. ²CNRS, iBV, UMR 7277, 06100 Nice, France. ³INSERM, iBV, U1091, 06100 Nice, France.

*These authors contributed equally to this work.

Data Fig. 1e–h). These two phenotypes were rescued by RNAi targeting the only known regulator of Hh secretion, Dispatched¹⁷ (Disp) (Fig. 1a and Extended Data Fig. 1e, i).

Expression in Hh-producing cells of RNAi against ESCRT proteins, including AliX, TSG101, Vps22 (also known as Larsen), Chmp1 and Chmp3 (members of the ESCRT complexes I, II and III; see Supplementary Table 1) abolished the Hh-dependent v3–v4 intervein space enlargement and WID anterior outgrowth (up to 48% of the WID showed complete rescue) (Fig. 1a and Extended Data Fig. 1e, j–k) without changing the level of Hh protein (Extended Data Fig. 1m, n). Rescued discs with wild-type appearance did not display any sign of apoptosis or cell-architecture defects, with the exception of Vps22, investigation of which was not further pursued (Extended Data Fig. 2a–k). Distribution of DE-cadherin (also known as Shotgun), Disp, and the glypican Dally-like (Dlp) in cells depleted for ESCRT function were also not affected (Extended Data Fig. 2h–s), indicating that the ESCRT depletion conditions used did not induce pleiotropic or indirect effects on Hh secretion.

The rescue of WID morphology revealed a role of the ESCRT proteins on the regulation of expression of the Hh long-range target *dpp*. Indeed,

very anterior ectopic expression of *dpp* and the broadening of its central stripe at the anterior–posterior border were abolished (Fig. 1d, e). Surprisingly, the expression of short-range targets requiring high-level signalling, such as En and Ptc, was still extended over up to six cells (Extended Data Fig. 3a–l). Consistent with a role of ESCRT in the long-range activity of Hh, accumulations of Hh at the external apical side of Hh-producing cells were observed upon depletion of ESCRT in posterior cells whereas the extracellular apical Hh levels in the anterior receiving cells were decreased compared with wild-type discs (Fig. 1f–o).

The requirement of ESCRT function for Hh long-range activity prompted us to test whether ESCRT function was also required in animals in which Hh is expressed at physiological levels. We thus depleted Chmp1, AliX, TSG101 and Vps32 (also known as Shrub) by RNAi (Supplementary Table 1) in the posterior cells of an otherwise wild-type disc. We observed a significant decrease of *dpp* and *iro* expression, whereas the domains of Ptc and En expression were similar to those in the wild type (Fig. 2a–e, v, x and Extended Data Fig. 4a–j, k). We obtained similar results by using the temperature-sensitive allele of the *hh* gene (*hh^{ts2}*), which increases the sensitivity for defects in Hh secretion (Extended Data Figs 1l, 4l–t).

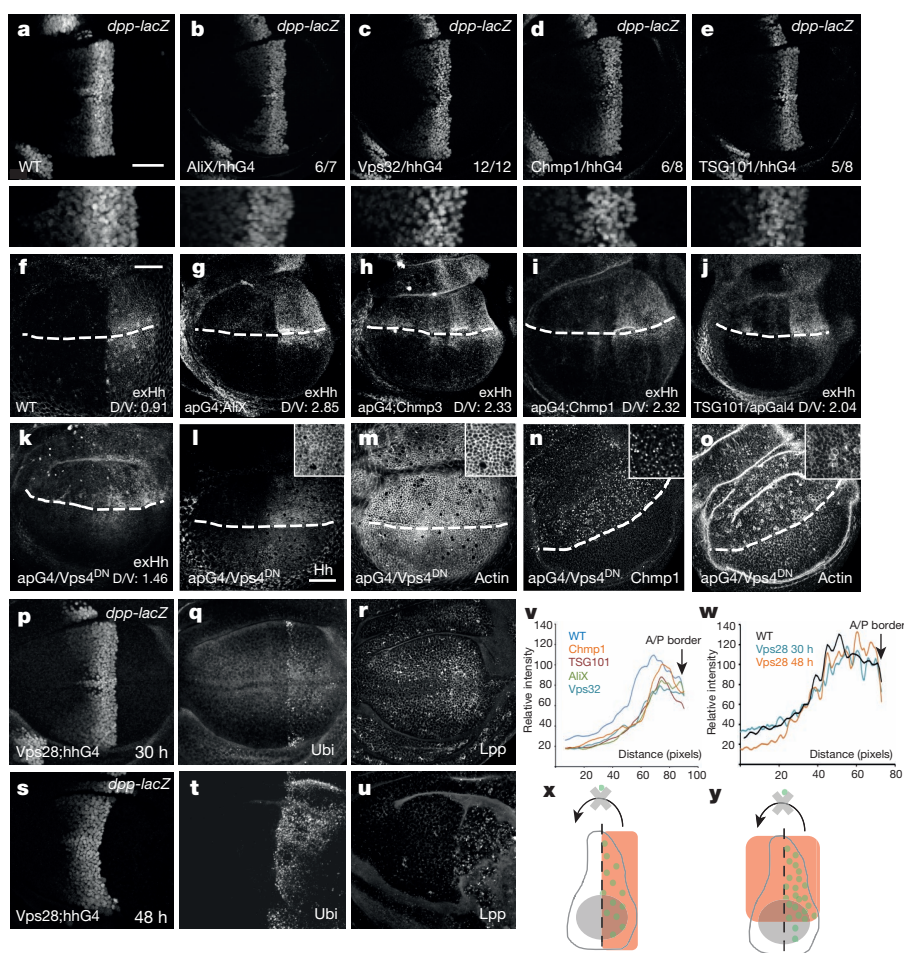


Figure 2 | Lack of ESCRT function in posterior cells leads to defects in Hh secretion and target gene induction. **a–e**, Pattern of *dpp-lacZ* expression in WIDs with 30 h of RNAi against the indicated ESCRT components. A magnification of the *dpp-lacZ* stripe is shown in the bottom panels. **b–e**, Numbers at bottom right indicate number of discs displaying reduced *dpp* activity per whole population analysed. **f–j**, Extracellular staining of endogenous Hh in wild-type (WT) discs and in discs with the indicated ESCRT RNAi. Dorsal is to the top and anterior to the left. **f–k**, The dorsal–ventral (D/V) ratio of Hh staining intensity is indicated in the bottom-right corner for all genotypes. **k–m**, Extracellular (exHh; **k**) and conventional (**l–m**) Hh staining in discs expressing Vps4^{DN}. **n, o**, Chmp1 accumulates in endosomal

compartments. F-actin staining is used to outline the cell cortex (**m, o**). Magnified views of dorsal–posterior cells are shown in the top-right corner. **f–o**, The dorsal–ventral boundary is marked with a white dashed line. **p–u**, Misexpression of Vps28 RNAi for 30 h (**p–r**) or 48 h (**s–u**) in Hh-producing cells. **v**, Quantification of *dpp-lacZ* from **a–e**. A/P, anterior–posterior. **w**, Quantification of *dpp-lacZ* expression in Vps28 RNAi discs presented in **p, s, x, y**. Schemes presenting the expression domain (red) of the compartment-specific Gal4 driver lines hedgehog-Gal4 (hhGal4; **x**) and apterous-Gal4 (apGal4; **y**). Green circles indicate Hh protein. Scale bars, 30 μ m.

To test whether impaired Hh activity was associated with defects of Hh subcellular localization, we depleted ESCRT proteins in the dorsal compartment of the WID (Fig. 2y). This led to the accumulation of Hh in dorsal ESCRT-depleted cells compared with their ventral wild-type neighbours. Accumulating Hh can be detected using a non-permeabilizing antibody staining protocol (see Methods), suggesting that the molecule is retained at the outer plasma membrane surface of producing cells (Fig. 2f–j). To confirm this, we induced the expression of a dominant-negative form of the ESCRT III protein Vps4 (Vps4^{DN}) to block the disassembly and recycling of the ESCRT complex^{18–20} in Hh-secreting cells. Surprisingly, intracellular Hh was not trapped in maturation-deficient endosomes labelled with Chmp1, but instead accumulated at the apical cortex of Hh-producing cells (Fig. 2k–o). This suggests that the ESCRT-dependent regulation of Hh secretion might not be related to the function of ESCRT proteins in MVB biogenesis. To confirm this, we analysed the effect of depletion of ESCRT function on MVB biogenesis. Blocking MVB function by expression of ESCRT RNAi or of Vps4^{DN} in Hh-producing cells led to the accumulation of poly-ubiquitinated epitopes (poly-Ubi) on proteins trafficking through degradative MVBs²¹ (Extended Data Fig. 5a–e, n–p). Interestingly, upon expression of Vps28 RNAi, this enrichment was not associated with a decrease in *dpp* target gene expression (Fig. 2p–q, s–t, w). We also showed that an increase in poly-Ubi accumulation in Vps32-depleted discs does not correlate with a decrease in *dpp-lacZ* expression (Extended Data Fig. 5m–p). Conversely, in AliX- and Vps32-depleted discs showing no accumulation of ubiquitinated proteins, expression of the long-range target *dpp* was reduced (Fig. 2b, c, v and Extended Data Fig. 5j, m). This indicates that defects in MVB trafficking do not correlate with defects in Hh signalling, but suggests that a subset of ESCRT proteins regulate Hh secretion in a manner that is independent of MVB biogenesis.

Hh can be associated with the Lipophorin (Lpp) lipid carrier (which labels lipoprotein particles) secreted from the fat body⁵; therefore, we wondered whether trafficking of Lpp is affected in the absence of ESCRTs. We found that depletion of Vps28—which has no effect on Hh signalling (Fig. 2p, s, w)—affects Lpp trafficking (Fig. 2r, u), whereas depletion of AliX and Vps32—which affects Hh secretion and signalling (Fig. 2b, c, g)—did not perturb Lpp trafficking (Extended Data Fig. 5k, l). In accordance with the fact that ESCRT-dependent Hh secretion is independent of Lpp, only the knockdown of ESCRT proteins but not the knockdown of Lpp led to the accumulation of Hh at the apical pole of the cells (Fig. 2g–l and Extended Data Fig. 5q–t). These data suggest that perturbation of the ESCRT machinery impairs Hh trafficking and secretion independently of Lpp.

To investigate whether secretion of Hh relies on the ESCRT-dependent biogenesis of exovesicle carriers, we used a cell line (Cl8 cells) derived from the WID and modified to express Hh in an inducible manner²². Biochemical fractionation by differential centrifugation of the Hh-containing conditioned medium was carried out (Methods and Fig. 3a). Western blot analysis indicated that the bulk of the Hh protein present in the conditioned medium was split equally between a 120,000g supernatant (S120) and a detergent-soluble pellet (P120) (Fig. 3a). Higher-speed centrifugation for a longer time gave similar fractionation (Extended Data Fig. 6a). The S120 fraction corresponds to a soluble Hh complex with a molecular weight of 160–200 kDa (ref. 22). The analysis by transmission electron microscopy of the P120 fraction from Hh-expressing Cl8 cells or of P120 from Cl8 cells not expressing Hh showed cup-shaped vesicles of 40–200 nm in diameter (Fig. 3b). Incubation of Hh-responsive Cl8 naive cells with an equivalent relative amount of P120 Hh or S120 Hh stimulated phosphorylation of the Fused protein—which reflects Hh signalling activation¹—to similar extents (Fig. 3c and Extended Data Fig. 6a, b), suggesting that these two fractions correspond to two different pools of Hh with similar capacities to induce Hh signalling. P120 Hh was also confined to a 35% (w/v) Optiprep step in a density gradient, corresponding to an intermediate density (between 1.184 and 1.195 g ml⁻¹), indicating specific membrane association (Fig. 3d). Anti-Hh staining by immunoelectron microscopy on non-permeabilized purified P120 membranes

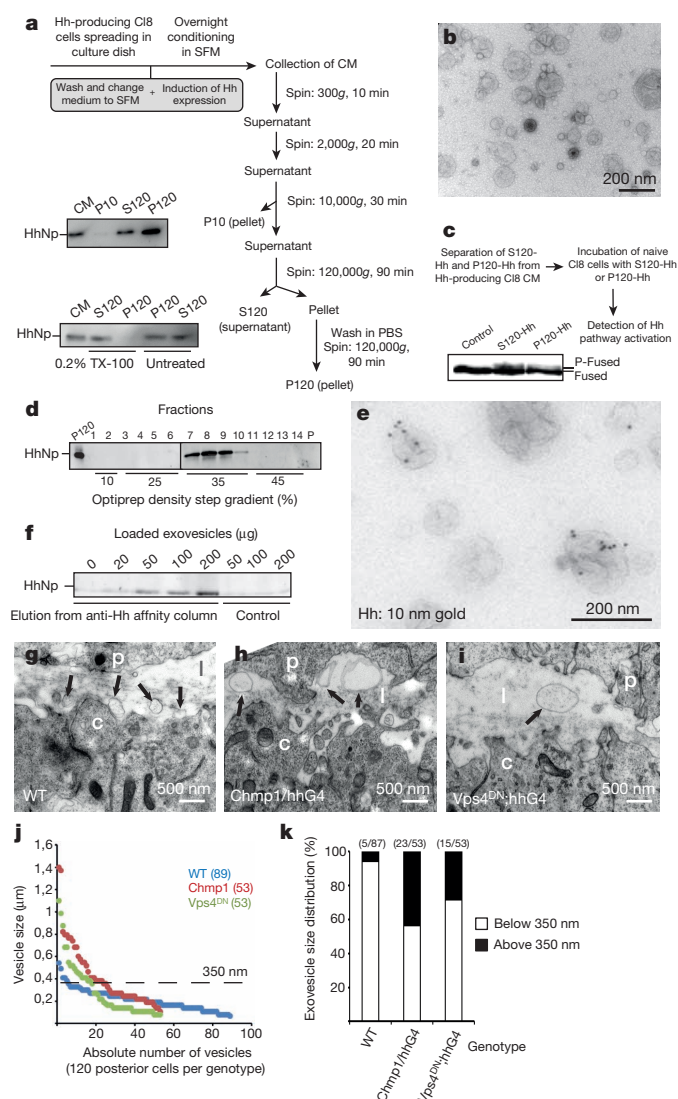


Figure 3 | Hh is secreted on extracellular vesicles. **a**, Conditioned medium (CM) from Hh-producing Cl8 cells subjected to differential centrifugation. The 10,000g pellet (P10), 120,000g pellet (P120) and 120,000g supernatant (S120) fractions were analysed by immunoblotting with anti-Hh antibodies (middle). Bottom, the Hh-producing Cl8-cell S120 and P120 fractions were treated with 0.2% Triton X-100 (TX-100). HhNp, mature cholesterol-modified HhN peptide; SFM, serum-free medium. **b**, Representative transmission electron micrograph of the Hh-producing Cl8-cell P120 fraction contrast stained with 3% uranyl acetate. **c**, Naive Cl8 cells were treated with Hh-producing Cl8-cell S120 or P120 fractions or with conditioned medium from Cl8 cells (control). Cell lysates from treated cells were analysed by immunoblotting with anti-Fused antibodies. P-Fused, phosphorylated Fused. **d**, Hh-producing Cl8-cell P120 was fractionated on an Optiprep density step gradient from 45% to 10%. **e**, Immunogold labelling of Hh P120 purified fraction with anti-Hh antibodies. **f**, Anti-Hh or control antibodies bound to Dynabeads were incubated with increasing amounts of Hh-producing Cl8-cell exovesicles from P120. Bound vesicles were eluted and analysed by immunoblotting with anti-Hh antibodies. **g–i**, Transmission electron micrographs of WIDs expressing Chmp1 RNAi and Vps4^{DN} in Hh-producing cells. White letters mark the apical side of peripodial (p) and columnar (c) cells facing the lumen (l). Black arrows indicate the luminal vesicular profiles. Scale bar, 500 nm. **j**, **k**, Quantification (**j**) and distribution (**k**) based on the size of vesicular profiles in Chmp1 RNAi and Vps4^{DN} compared with wild type (WT). Numbers in brackets indicate the total number of vesicular profiles counted (**j**), or the number of vesicular profiles sized above 350 nm per whole population (**k**).

revealed that Hh was associated with the surface of vesicles (Fig. 3e) but not on vesicles purified from non-Hh-expressing cells (Extended Data Fig. 6c). Moreover, when Hh-containing vesicles were loaded onto a Hh immunofluorescence column we found that Hh was detectable in fractions eluted from the column, confirming the presence of Hh epitopes on the surface of vesicles (Fig. 3f).

We next identified the proteins present in the Hh-containing vesicles by performing mass spectrometry on the P120 Hh fraction. We identified about 160 proteins, including the Hh amino-terminal peptide, which contains all the signalling activity of Hh (Supplementary Table 2). These proteins included nine ESCRT protein members, five of them (AliX, Chmp1, Chmp3, Vps32 and Vps4) being the ones that affect Hh secretion *in vivo*.

We further investigated ESCRT function using electron microscopy. Careful examination of the plasma membrane and luminal space between the columnar and peripodial cells revealed the presence of extracellular vesicular profiles in the luminal space, while MVB morphology was not significantly affected (Fig. 3g–k and data not shown). Interestingly, these structures are bigger in size and their number is decreased in Chmp1 RNAi- and in Vps4^{DN}-expressing discs (Fig. 3g–k). The accumulation of Hh at the surface of producing cells (Fig. 2g–l), combined with the results of our electron microscopy study, suggest that, as for exovesicle release from HEK293T cells⁸ and in *Caenorhabditis elegans*⁹, Hh may be released through plasma membrane budding into the extracellular space. We therefore analysed the *in vivo* association of extracellular Hh with ESCRT proteins in more detail. In WIDs expressing green fluorescent protein (GFP)-tagged Vps32 in Hh-producing cells we observed dynamic Vps32-GFP-positive particles in the luminal space between columnar epithelial and squamous peripodial cells, at a distance from producing cells (Fig. 4a, b and Extended Data Fig. 7a, b). Interestingly, extracellular Hh partly colocalized with Vps32-GFP, at the apical external side of both posterior producing and anterior receiving Hh cells (Fig. 4c–e and Extended Data Fig. 8a–i). No such colocalization was observed between external Hh and an inner plasma membrane marker (GAP43-GFP) (Extended Data Fig. 8j–l). Moreover, endogenous Chmp1-positive external particles were observed at various places in the luminal space (Fig. 4f–h and Extended Data Figs 7c–f, 9). Importantly, extracellular Hh-GFP was found to be colocalized with endogenous Chmp1 in the luminal space (Fig. 4f–h and Extended Data Fig. 7c–f).

It has been previously shown that Hh is present in the fly haemolymph²³, which we therefore used as another *in vivo* model to analyse Hh secretion. In this tissue, we found similar overlap between Vps32-GFP and endogenous Hh on secreted particles (Fig. 4i–k). Such overlap was also found with endogenous AliX (for validation of the anti-AliX antibody see Extended Data Fig. 5i), but not with GAP43-GFP or Vps36-GFP (Extended Data Fig. 10), suggesting that, in this tissue also, Hh is associated with particles that are positive for a subset of ESCRT proteins.

If Hh is secreted together with ESCRT proteins, then misexpression of the Hh-receptor Ptc should trap both Hh and ESCRT mixtures at the cell surface. We investigated this possibility by expressing a mutant form of Ptc, Ptc(1130X), in the Hh-receiving cells at the anterior–posterior border. Ptc(1130X) can bind and sequester external Hh but cannot be internalized. Consistent with a potential role of Chmp1-positive exovesicles in the delivery of Hh to its target tissue, Ptc(1130X) expression led to the trapping of endogenous Chmp1 and Hh at the surface of recipient cells in the anterior compartment (Fig. 4l–n). By contrast, Ptc(1130X) misexpression did not cause any detectable trapping of Lpp (Fig. 4o–q), indicating that ESCRT/Hh-positive exovesicles are distinct from previously described Lpp-positive Hh carriers⁵.

Our findings provide evidence for a previously unidentified mechanism for Hh release and intercellular communication (see Supplementary Discussion). Overall, our results suggest a new mechanism for morphogen secretion, in which a pool of signalling molecules follows an ESCRT-dependent route in morphogen-producing cells, contributing to the establishment of a morphogen gradient. We believe that this mechanism

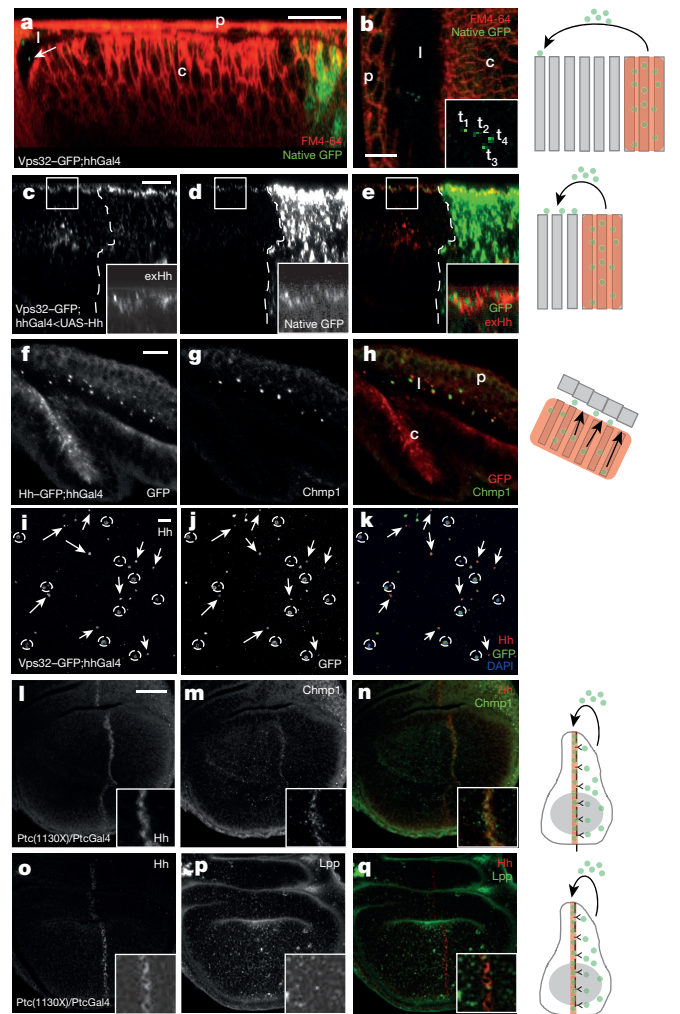


Figure 4 | ESCRT-positive exovesicles transport Hh in the extracellular space of *Drosophila* wing discs and haemolymph. **a**, X–Z section of a living WID in which Vps32–GFP is expressed in the Hh-secreting posterior compartment (green cells on the right). The arrow indicates a Vps32-positive particle in the lumen. Scale bar, 20 μ m. Plasma membrane was labelled with the vital membrane dye FM4-64 [(N-(3-(triethylammoniumpropyl)-4-(6-(4-(diethylamino) phenyl) hexatrienyl) pyridinium dibromide)]. **b**, Temporal projection of a time-lapse video of a Vps32–GFP particle moving through the luminal space. Single time points (acquisition time: 0.77 s per frame) are indicated in the bottom right corner (t1–t4). Scale bar, 10 μ m. **c–e**, X–Z section of a disc showing partial colocalization of extracellular Hh (exHh) and Vps32–GFP in Hh-receiving cells. Apical is to the top, anterior to the left. The anterior–posterior border is labelled with a white dashed line. Scale bar, 10 μ m. Pearson coefficient: 0.576; M1M2 coefficients: 0.308 (Hh) and 0.593 (Vps32–GFP). **f–h**, Colocalization of Hh–GFP and endogenous Chmp1 protein in particles in the luminal space (see also Extended Data Fig. 7). Scale bar, 10 μ m. **i–k**, Colocalization of Hh and Vps32–GFP in non-cellular particles within the haemolymph. White circles mark 4',6'-diamidino-2-phenylindole (DAPI)-positive blood cells, white arrows indicate examples of colocalization. Scale bar, 20 μ m. **l–q**, Trapping of Hh and Chmp1 but not Lpp at Ptc(1130X)-expressing cells. Scale bar, 20 μ m. **c**, columnar cells; **l**, lumen; **p**, peripodial cells. Magnification is shown in the bottom-right corner of **c–e**, **l–n** and **o–q**. Schemes on the right column indicate position of columnar cells (in red) that express the drivers used in **a–q**. Green circles indicate Hh, and black branches indicate Ptc(1130X) proteins.

of Hh secretion may be conserved, as the vertebrate counterpart of Hh, Sonic hedgehog (Shh), has been observed on exovesicles in the mouse ventral node²⁴. However, it remains to be determined whether mammalian ESCRTs are involved in this process and whether functionally active vesicular Shh spreads through the extracellular space.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 September 2013; accepted 8 September 2014.

1. Briscoe, J. & Théron, P. P. The mechanisms of Hedgehog signalling and its roles in development and disease. *Nature Rev. Mol. Cell Biol.* **14**, 416–429 (2013).
2. Mann, R. K. & Beachy, P. A. Novel lipid modifications of secreted protein signals. *Annu. Rev. Biochem.* **73**, 891–923 (2004).
3. Rusten, T. E., Vaccari, T. & Stenmark, H. Shaping development with ESCRTs. *Nature Cell Biol.* **14**, 38–45 (2012).
4. Zeng, X. *et al.* A freely diffusible form of Sonic hedgehog mediates long-range signalling. *Nature* **411**, 716–720 (2001).
5. Panáková, D., Sprong, H., Marois, E., Thiele, C. & Eaton, S. Lipoprotein particles are required for Hedgehog and Wingless signalling. *Nature* **435**, 58–65 (2005).
6. Sanders, T. A., Lagostera, E. & Barna, M. Specialized filopodia direct long-range transport of SHH during vertebrate tissue patterning. *Nature* **497**, 628–632 (2013).
7. Bischoff, M. *et al.* Cytonemes are required for the establishment of a normal Hedgehog morphogen gradient in *Drosophila* epithelia. *Nature Cell Biol.* **15**, 1269–1281 (2013).
8. Nabhan, J. F., Hu, R., Oh, R. S., Cohen, S. N. & Lu, Q. Formation and release of arrestin domain-containing protein 1-mediated microvesicles (ARMMs) at plasma membrane by recruitment of TSG101 protein. *Proc. Natl Acad. Sci. USA* **109**, 4146–4151 (2012).
9. Wehman, A. M., Poggioli, C., Schweinsberg, P., Grant, B. D. & Nance, J. The P4-ATPase TAT-5 inhibits the budding of extracellular vesicles in *C. elegans* embryos. *Curr. Biol.* **21**, 1951–1959 (2011).
10. Raposo, G. & Stoorvogel, W. Extracellular vesicles: exosomes, microvesicles, and friends. *J. Cell Biol.* **200**, 373–383 (2013).
11. Torroja, C., Gorfinkel, N. & Guerrero, I. Mechanisms of Hedgehog gradient formation and interpretation. *J. Neurobiol.* **64**, 334–356 (2005).
12. Gómez-Skarmeta, J. L. & Modolell, J. *araucan* and *caupolican* provide a link between compartment subdivisions and patterning of sensory organs and veins in the *Drosophila* wing. *Genes Dev.* **10**, 2935–2945 (1996).
13. Vaccari, T. & Bilder, D. The *Drosophila* tumor suppressor *vps25* prevents nonautonomous overproliferation by regulating notch trafficking. *Dev. Cell* **9**, 687–698 (2005).
14. Herz, H.-M. *et al.* *vps25* mosaics display non-autonomous cell survival and overgrowth, and autonomous apoptosis. *Development* **133**, 1871–1880 (2006).
15. Thompson, B. J. *et al.* Tumor suppressor properties of the ESCRT-II complex component *Vps25* in *Drosophila*. *Dev. Cell* **9**, 711–720 (2005).
16. Vaccari, T. *et al.* Comparative analysis of ESCRT-I, ESCRT-II and ESCRT-III function in *Drosophila* by efficient isolation of ESCRT mutants. *J. Cell Sci.* **122**, 2413–2423 (2009).
17. Burke, R. *et al.* Dispatched, a novel sterol-sensing domain protein dedicated to the release of cholesterol-modified hedgehog from signaling cells. *Cell* **99**, 803–815 (1999).
18. Babst, M., Sato, T. K., Banta, L. M. & Emr, S. D. Endosomal transport function in yeast requires a novel AAA-type ATPase, *Vps4p*. *EMBO J.* **16**, 1820–1831 (1997).
19. Fujita, H. *et al.* A dominant negative form of the AAA ATPase SKD1/VPS4 impairs membrane trafficking out of endosomal/lysosomal compartments: class E *vps* phenotype in mammalian cells. *J. Cell Sci.* **116**, 401–414 (2003).
20. Rusten, T. E. *et al.* ESCRTs and *Fab1* regulate distinct steps of autophagy. *Curr. Biol.* **17**, 1817–1825 (2007).
21. Raiborg, C. & Stenmark, H. The ESCRT machinery in endosomal sorting of ubiquitinated membrane proteins. *Nature* **458**, 445–452 (2009).
22. Gallet, A., Ruel, L., Staccini-Lavenant, L. & Théron, P. P. Cholesterol modification is necessary for controlled planar long-range activity of Hedgehog in *Drosophila* epithelia. *Development* **133**, 407–418 (2006).
23. Palm, W. *et al.* Secretion and signaling activities of lipoprotein-associated Hedgehog and non-sterol-modified Hedgehog in flies and mammals. *PLoS Biol.* **11**, e1001505 (2013).
24. Tanaka, Y., Okada, Y. & Hirokawa, N. FGF-induced vesicular release of Sonic hedgehog and retinoic acid in leftward nodal flow is critical for left–right determination. *Nature* **435**, 172–177 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the French Government (National Research Agency, ANR) through the “Investments for the Future” LABEX SIGNALIFE (program reference number ANR-11-LABX-0028-01) and by the Fondation pour la Recherche Médicale (reference number DEQ20110421324). T.M. and F.W. were supported by the Fondation Association pour la Recherche Contre le Cancer (ARC) and Ligue Nationale Contre le Cancer, respectively. M.F. was supported by ATIP/Avenir, ARC and the Human Frontier Science Program. The authors acknowledge S. Pagnotta for technical support. We thank K. Haglund for sharing the anti-AliX antibody before publication. We thank L. Staccini-Lavenant and M.-A. Derieppe for technical help.

Author Contributions P.P.T. conceived the project and supervised the study. P.P.T. and T.M. designed the genetic studies. T.M. performed all the genetic tests and most of the immunohistochemistry on WIDs. F.W. and P.P.T. designed the biochemical studies. F.W. performed the biochemistry/molecular biology experiments. M.F. designed and performed the live imaging experiments and generated the Chmp1b antibody. M.F. and S.Po. analysed the localization of Hh, Chmp1 and Lpp in wild-type animals and in discs expressing *Vps4^{DN}*, *Ptc(1130X)* and *Lpp RNAi*. G.D.A. produced reagents and contributed to the experimental design of extracellular labelling of Hh. S.Pi. performed the electron microscopy and immunoelectron microscopy analysis, and performed the counting on the electron microscopy data. P.P.T., M.F., F.W., S.Pi. and T.M. analysed the individual collected data sets. P.P.T. and T.M. assembled and edited the figures. P.P.T. wrote the manuscript. All authors commented on the manuscript versions. S.Pi. and S.Po. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.P.T. (therond@unice.fr).

METHODS

Fly stocks and genetics. Fly stocks were grown under standard conditions. For investigations of the effect of ESCRT RNAi in Hh gain-of-function conditions, the *w;tubGal80^{ts};hhGal4 < UAS-Hh/TM6b,Tb* 'tester line' was crossed with either control or ESCRT RNAi lines. Starter crosses were flipped every 12 h at 25 °C, to synchronize the progeny, and vials with L2 instar larvae were transferred to 29 °C. The induction of upstream activating sequence (UAS)-Hh expression in this setup leads to strong pharate adult lethality, the amount of eclosed escaper flies in the control *w;tubGal80^{ts}/+; hhGal4 < UAS-Hh/+* genotype is below 1%. For the WID morphology assay, non-Tb larvae were dissected at the L3 larval instar stage. As a control for suppression of Hh-dependent outgrowth, we used the *Disp^{10004GD}* RNAi line, which gives no phenotype in otherwise wild-type animals (data not shown), confirming the greater sensitivity of our experimental set-up for defects in Hh secretion. In Hh loss-of-function experiments we used the *hhGal4* (enhancer trap for the *hh* locus, weakly hypomorphic by itself)/*hh^{ts2}* (conditional null allele²⁵) heteroallelic combination, which is viable at permissive temperature (18 °C) and semi-lethal (up to 60% of the flies die as pharate adults) after incubation at restrictive temperature (29 °C) for 3 days. For simplicity, we refer in the text to *hh^{ts2}/hhGal4*. ESCRT RNAi; *hh^{ts2}/SM6bTM6b,Tb* males were crossed with *w;dpp-nuc-lacZ^{P10638}*, *hhGal4/SM6bTM6b,Tb* virgin females at 18 °C. Crosses were kept at 18 °C and flipped daily. Vials containing L1 larvae were shifted to 29 °C, and non-Tb L3 larvae were dissected after 4 days of induction. For visualization of the changes in the levels of endogenous extracellular Hh, wild-type control or ESCRT RNAi lines were crossed with *apGal4 < dpp-nuc-lacZ^{P10638}*; *MKRS/SM6bTM6b,Tb*. Crosses were kept at 25 °C, flipped every 12 h, and non-Tb L3 larvae from the progeny were dissected for further analysis.

The RNAi lines used in this study were: *Ci^{105620KK}* (VDR), *Disp^{10004GD}* (VDR). Detailed description about the usage of ESCRT RNAi-s and further ESCRT properties can be found in Supplementary Table 1. The other alleles used were *w¹¹¹⁸* as a wild-type control; *hh^{ts2}* (Bloomington #1684), *iro-lacZ²⁶*, the nuclear *dpp-nuc-lacZ^{P10638}* reporter (provided by L. Xinhua), *UAS-Ptc^{1130X}* (Bloomington #44612)²⁷.

Live imaging of Vps32-GFP in Drosophila WIDs. For visualization of the distribution of Vps32-GFP in living WIDs, animals of the *UAS-Vps32-GFP/+; hhGal4 < tubGal80^{ts}/+* genotype were kept at room temperature until the late third larval instar stage and were then shifted to 29 °C the evening (or for 6 h when looking at both Hh and Vps32-GFP distribution in fixed discs) before dissection.

Dissected WIDs were mounted in *Drosophila* Shields & Sang M3 medium (Sigma) supplemented with FM4-64 (1:2,000 from 10 µg µl⁻¹ stock, Invitrogen F34653) and 0.7% low melting point agarose.

Adult wing preparation. Adult flies were collected in glycerol:ethanol (3:1) and stored for at least 24 h before dissection. Adult wings were removed with forceps at their hinge and briefly rinsed in 96% ethanol before mounting in Euparal, as previously described²⁸. Preparations were kept at 60 °C on a slide-heating plate overnight before image acquisition. Bright-field images were taken with a Leica Axioplan2 microscope (objective, ×5; Plan NEOFLUAR; NA = 0.15) equipped with a Leica DFC490 digital camera. For image acquisition, we used LAS v.2.5.0 R1 (build 975).

Dissections and immunohistochemistry. Conventional staining was performed as described previously²⁹. For imaging, WIDs were mounted in VECTASHIELD mounting medium (Vector Laboratories).

For extracellular Hh and GFP staining, L3 larvae were collected and cleaned in ice-cold PBS. Larvae were then dissected in ice-cold Schneider's medium (Sigma) and incubated with primary antibodies in Schneider's medium at 4 °C for 1 h. Following this step, samples were rinsed three times with ice-cold PBS and fixed for 15 min in PBS plus 2% formaldehyde at 4 °C. Fixed tissue then was washed with PBS (three times for 10 min each in PBS at room temperature) and blocked by incubation for 30 min in PBS plus 2% bovine serum albumin (BSA). Secondary antibodies in PBS plus 2% BSA were added and the samples were incubated for 1 h at room temperature and then washed in PBS and mounted as described earlier.

For the visualization of exovesicles containing Hh-GFP and endogenous Chmp1 protein in WIDs, animals of the *hhGal4 < tubGal80^{ts}/UAS-Hh-GFP* genotype were kept at room temperature until the late third larval instar stage and were then shifted to 29 °C on the evening before dissection.

For analysis of the effect of *Vps4^{DN}* on the localization of Hh and Chmp1, animals of the *UAS-Vps4^{DN}/tubGal80^{ts};hhGal4 < UAS-Hh/+* and *UAS-Vps4^{DN}/tubGal80^{ts}; hhGal4/+* genotypes were kept at 18 °C until the late third larval instar stage and were then shifted to 29 °C for the last 8 h before dissection.

We induced *Ptc*(1130X) expression in Hh-receiving cells, by keeping animals of the *ptcGal4 < tubGal80^{ts}/UAS-ptc^{1130X}* genotype at room temperature for 24 h after egg laying, after which they were shifted to 29 °C until dissection.

For haemolymph immunohistochemistry, ten larvae were bled on ice for 1 min and 1 µl of the haemolymph was dispersed on poly-L-lysine-coated 8-well tissue chamber slides (MAT-TEK). The preparation was then dried in vacuum for 5 min and immediately fixed in PBS plus 2% formaldehyde for 20 min at room temperature.

The fixed haemolymph was then washed with PBS three times for 5 min and incubated with PBS plus 5% FCS for 1 h. Following this step, slides were incubated with primary antibodies overnight at 4 °C, and afterwards washed three times for 10 min with PBS. Secondary antibodies were diluted in PBS plus 5% FCS and added to the slides for 1 h. Finally, before mounting in VECTASHIELD supplemented with DAPI, slides were washed again three times for 10 min with PBS.

Mounted WID preparations and haemolymph were imaged with a Leica TCS SP5 confocal microscope (objective, ×40; HCX PLAN APO; NA = 1.3) and the LAS AF 2.6.3.8173 program, a Zeiss LSM510 (objective, ×40; PLAN APO; NA = 1.3) or a ZEISS LSM780 (objective, ×63; PLAN APO; NA = 1.4). Bright-field WID images were taken with a Leica Axioplan2 microscope (objective, ×10; Plan NEOFLUAR; NA = 0.3).

Antibodies used in this study. Primary antibodies used were as follows. Rabbit anti-Hh³⁰ (1:500 from crude serum for conventional, 1:100 from purified serum for extracellular), rat anti-Hh (1:250, this study), mouse anti-GFP (1:1,000 for conventional, 1:50 for extracellular; Roche #1814460), mouse anti-Dlg (1:50; DSHB 4F3), rat anti-DE-cadherin (1:50; DSHB DCAD2), chicken anti-β-gal (1:1,000; GeneTEX GTX77365), rat anti-Ci 2A1 (provided by B. Holmgren; 1:10), mouse anti-Ptc 5E10 (provided by P. Ingham; 1:400), rabbit anti-En (1:1,000; Santa Cruz sc-28640), rabbit anti-Caspase III (1:500; Cell Signaling #9661), mouse-anti-Ubi (clone FK2; 1:1,000; ENZO Life Sciences, BML-PW8810-0100), guinea-pig-anti-AliX (1:2,000; provided by K. Haglund before publication), rabbit-anti-Lpp and guinea-pig-anti-Lpp (both 1:500; provided by S. Eaton).

Chmp1 protein was detected using a custom-made antibody (used at a dilution of 1:300) generated through immunization with an antigenic peptide (QDELSQLAKLRDQV) corresponding to a carboxy-terminal epitope of zebrafish Chmp1b that is largely conserved in *Drosophila melanogaster* (Eurogentec).

Secondary antibodies used were as follows. Goat anti-rabbit Alexa488 (1:500; Invitrogen A11034), goat anti-rabbit Alexa546 (1:500; Invitrogen A11035), goat anti-rabbit DyLight649 (1:500; Jackson ImmunoResearch Laboratories 111-605-003), goat anti-rabbit Cy5 (1:200; Jackson ImmunoResearch Laboratories 111-175-144), donkey anti-mouse Alexa488 (1:500; Invitrogen A21202), donkey anti-mouse Alexa546 (1:500; Invitrogen A10036), donkey anti-rat Alexa488 (Invitrogen A21208), goat anti-rat Alexa546 (1:500; Invitrogen A11081), goat anti-chicken Alexa488 (1:500; Invitrogen A11039) and TRITC-phalloidin (1:100; Sigma P1951-1MG) were used. **Image analysis.** Bright-field and confocal images were edited in Adobe Photoshop CS5 Extended (x64, v.12) or Fiji 1.48a. For intensity measurements, intensity profiles were originally determined with Fiji. Raw intensity values were then further processed in Microsoft Excel 2010. To calculate Pearson and M1M2 coefficients the Fiji JACoP plugin was used³¹.

Hh exovesicle isolation and partial purification. Flasks of confluent cells (from 20 to 60 75-cm² flasks) were washed the day before the experiment, with 10 ml PBS. They were then incubated overnight with 4 ml serum-free Schneider's medium supplemented with 100 µM CuSO₄ for Hh expression induction. Conditioned medium was then subjected to centrifugation at 300g for 10 min and the supernatant was then centrifuged a second time at 2,000g for 20 min. The supernatant was then cleared by centrifugation at 10,000g for 30 min and then at 120,000g (a single centrifugation at 160,000g gave similar results) for 90 min. The P120 pellet was resuspended in PBS and centrifuged at 120,000g for 90 min. All steps were performed at 4 °C. The resuspended pellet was then immediately frozen in 300–500 µl aliquots in liquid nitrogen, giving a typical protein concentration of 5–10 µg µl⁻¹, and stored for up to 3 months at –70 °C.

Typically, we loaded 25% of the P120 obtained from 1 ml of conditioned medium per lane. An equal amount of the S120 pellet was loaded after TCA precipitation and resuspension in 1 × loading buffer. A similar protocol was followed for the isolation of exovesicles from cells not producing Hh. When indicated, vesicle preparations were subjected to equilibrium centrifugation. In such cases, the resuspended P120 was adjusted to 400 µl of 45% (w/v) Optiprep (60% iodixanol stock; Axis-Shield) and layered on top of a step gradient consisting of 400 µl 35%, 400 µl 25% and 200 µl 10% Optiprep (w/v). After centrifugation for 16–18 h at 120,000g, 100 µl aliquots were collected, starting from the top of the tubes, and snap-frozen in liquid N₂. Immunoblotting was performed as previously described³².

Immunoelectron microscopy. Fractions (7 to 9 from the Optiprep gradient) purified from Hh-expressing or non-expressing cells were fixed in 2% paraformaldehyde, deposited on Formvar carbon-coated electron microscopy grids and immunolabelled with rabbit anti-Hh TK antibody (a gift from T. Kornberg) and protein A-gold conjugates, as described previously³³. Exovesicles were observed at 80 kV with a 120 kV JEOL JEM-1400 electron microscope and digital acquisitions were obtained with a MORADA digital camera (Olympus SIS). Protein A-gold conjugates (10 nm) were purchased from the Cell Microscopy Center, Department of Cell Biology, Utrecht University.

Transmission electron microscopy. WIDs were fixed in 1.5% glutaraldehyde in 0.075 M cacodylate buffer and embedded in Epon according to standard procedures.

Seventy-nanometre sections were observed at 100 kV with a 120 kV JEOL JEM-1400 electron microscope and digital acquisitions were obtained with a MORADA digital camera (Olympus SIS).

Both immuno- and transmission electron microscopy were performed on the Centre Commun de Microscopie Appliquée electron microscopy core facility (Université Nice Sophia Antipolis).

Immunoisolation. Dynabead (Life Technologies) magnetic beads already coupled to protein A were saturated with rabbit serum containing either anti-Hh or control IgGs. The prepared beads were then incubated with Hh vesicles from enriched fractions. The beads were washed according to the manufacturer's protocol and bound material was eluted with 0.1 mM glycine pH 2 and subjected to western blotting.

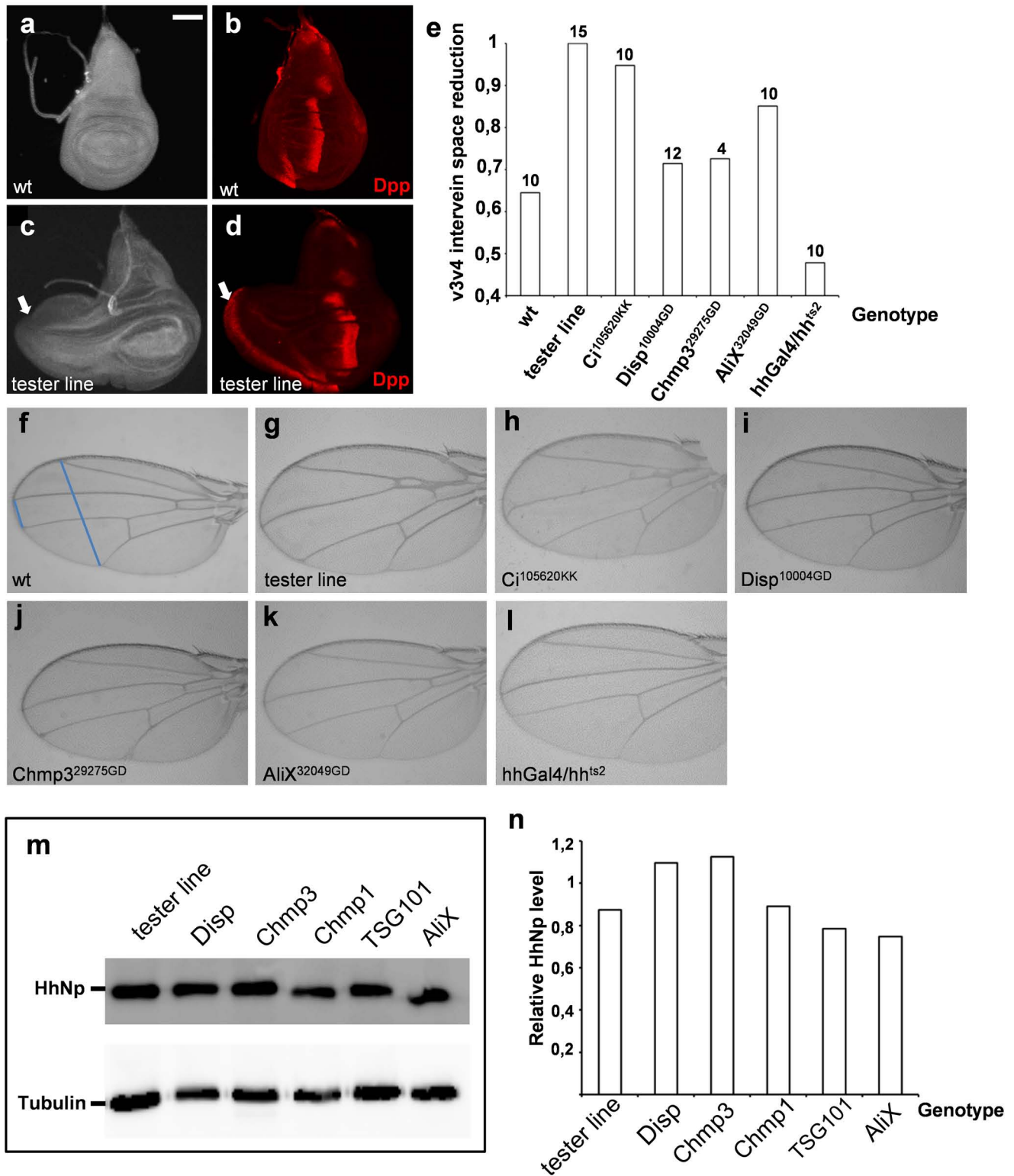
Hh response assay. On the day before the assay, Cl8 cells were used to seed six-well plates at 50% confluence. The following day, the cells were washed in PBS and incubated with P120 or S120 fractions (from the equivalent of 3 ml of Hh-producing Cl8-cell conditioned medium) for 1 day. Cells were then scraped off the plates, washed twice and lysed. The lysate was centrifuged at 12,000g and the supernatant was analysed by western blotting for Fused mobility shift (anti-Fu³⁴ 1:1,000). Equal relative amounts of P120 and S120 were analysed, to determine the amount of Hh before and after incubation with Cl8 cells, by western blotting (anti-Hh: 1:1,000; data not shown).

Western blot on rescued WIDs. One-hundred rescued discs of *ESCRT RNAi/tubGal80^{ts};hhGal4 < UAS-Hh/+* genotype or controls (wild type, *Disp^{10004GD}*) were collected into ice-cold PBS, and extracted in Laemmli buffer. The lysate was analysed on 12% SDS-PAGE and developed with anti-Hh and anti-Tubulin. Band density was measured with Fiji and raw data were analysed in Microsoft Excel.

Hh-P120 analysis by combined liquid chromatography with tandem mass spectrometry. Mass spectrometry analysis was performed on the Hh-containing P120 fraction obtained from Hh-producing Cl8-cell conditioned medium. Processed samples were sent to the European Molecular Biology Laboratory (EMBL) mass spectrometry core facility centre (http://www.embl.de/proteomics/proteomics_services) and were treated according to the facility protocol (detailed protocol can be provided upon request). In brief, peptides resulting from either in-gel or in-solution digestion

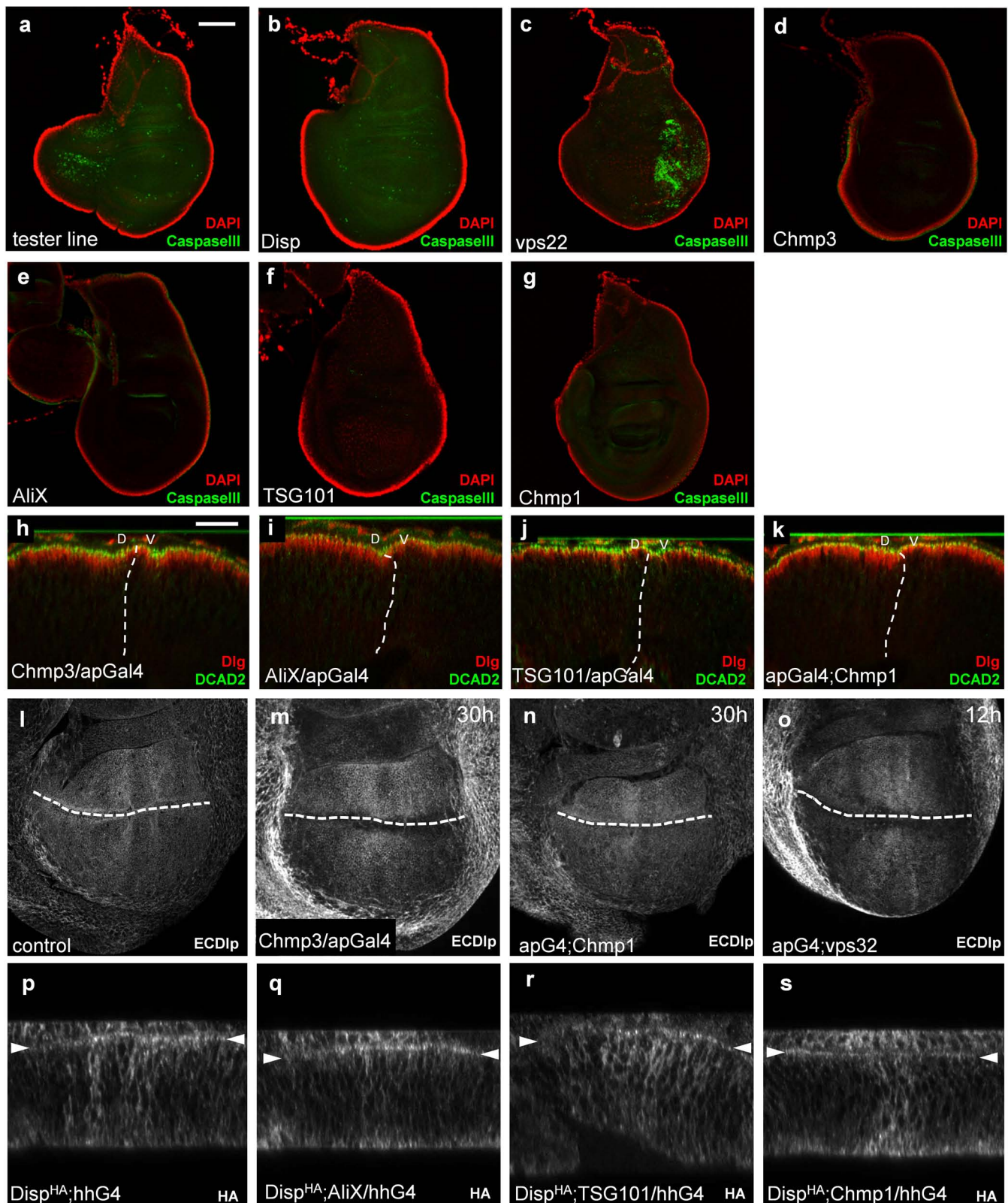
with trypsin were analysed by combined liquid chromatography with tandem mass spectrometry (LC-MS/MS) on a nano-high-performance liquid chromatography (HPLC) system (Proxeon) coupled directly to either an LTQ Orbitrap Velos (Thermo Fisher Scientific) or MaXis Q-ToF (Bruker). Data were processed using either MaxQuant (v.1.0.13.13) or within Hystar (Bruker) to generate files that could be submitted to Mascot (Matrix Science) for database searching against a UniProt species specific database (*D. melanogaster*).

25. Ma, C., Zhou, Y., Beachy, P. A. & Moses, K. The segment polarity gene hedgehog is required for progression of the morphogenetic furrow in the developing *Drosophila* eye. *Cell* **75**, 927–938 (1993).
26. Letizia, A., Barrio, R. & Campuzano, S. Antagonistic and cooperative actions of the EGFR and Dpp pathways on the *irquois* genes regulate *Drosophila* mesothorax specification and patterning. *Development* **134**, 1337–1346 (2007).
27. Johnson, R. L., Milenkovic, L. & Scott, M. P. *In vivo* functions of the patched protein: requirement of the C terminus for target gene inactivation but not Hedgehog sequestration. *Mol. Cell* **6**, 467–478 (2000).
28. Adler, P. N., Charlton, J. & Vinson, C. Allelic variation at the *frizzled* locus of *Drosophila*. *Dev. Genet.* **8**, 99–119 (1987).
29. Gallet, A., Ruel, L., Staccini-Lavenant, L. & Théron, P. P. Cholesterol modification is necessary for controlled planar long-range activity of Hedgehog in *Drosophila* epithelia. *Development* **133**, 407–418 (2006).
30. Gallet, A., Rodriguez, R., Ruel, L. & Théron, P. P. Cholesterol modification of Hedgehog is required for trafficking and movement, revealing an asymmetric cellular response to Hedgehog. *Dev. Cell* **4**, 191–204 (2003).
31. Bolte, S. & Cordelières, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).
32. Ruel, L., Rodriguez, R., Gallet, A., Lavenant-Staccini, L. & Théron, P. P. Stability and association of Smoothened, Costal2 and Fused with Cubitus interruptus are regulated by Hedgehog. *Nature Cell Biol.* **5**, 907–913 (2003).
33. Théry, C., Amigorena, S., Raposo, G. & Clayton, A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Curr. Protoc. Cell Biol.* Chapter 3, Unit 3.22 (2006).
34. Théron, P. P., Knight, J. D., Kornberg, T. B. & Bishop, J. M. Phosphorylation of the fused protein kinase in response to signaling from hedgehog. *Proc. Natl Acad. Sci. USA* **93**, 4224–4228 (1996).



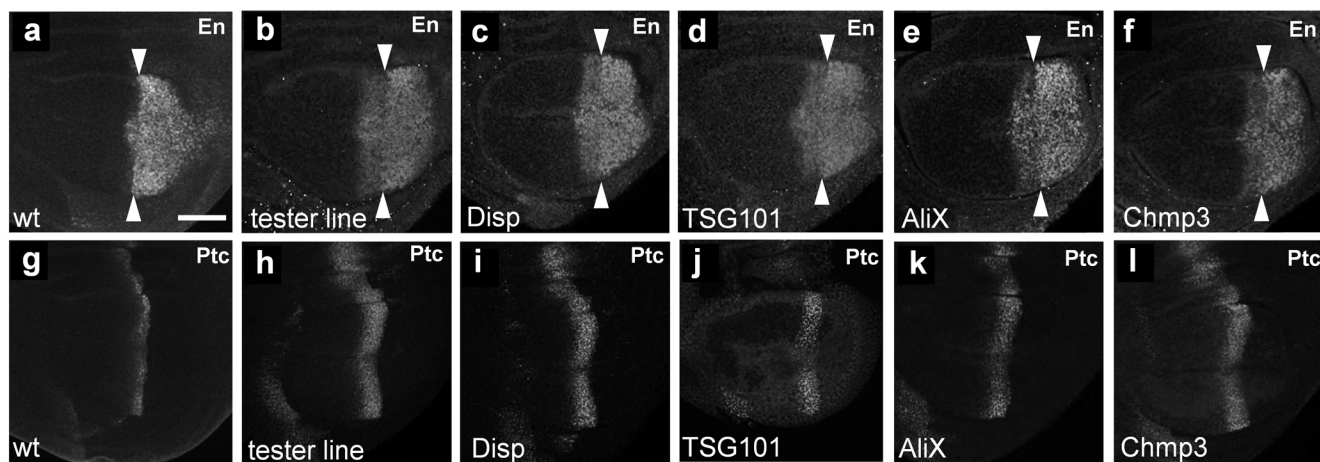
Extended Data Figure 1 | ESCRT depletion in posterior cells rescues the Hh gain-of-function phenotype in adult *Drosophila* wings. **a–d**, Morphology and *dpp-lacZ* expression pattern of wild-type (WT; **a**, **b**) and Hh-overexpressing WIDs (**c**, **d**; tester line). Scale bar, 100 μ m. Dorsal is towards the top and anterior is to the left. The white arrow in **c** indicates anterior outgrowth, whereas in **d** it points to cells expressing ectopic *dpp-lacZ*. **e**, Quantification (mean of the ratios of absolute values) of the v3–v4 intervein space (shorter blue line on **f**) relative to the whole wing span (longer blue line on **f**) normalized to the tester line. Sample numbers (indicating the number of wings examined)

are shown at the top of each column. **f–k**, Representative examples of controls (**f–i**) and wings of tester line plus ESCRT RNAi (**j**, **k**) escapers. The amount of eclosed escaper flies in the control *w;tubGal80^{ts}/+;hhGal4<UAS-Hh/+* genotype is below 1%. **l**, *hhGal4/hh^{ts2}* mutant wing from animals kept at restrictive temperature for 4 days from L1 larval stage. **m**, Representative western blot of WID lysates from rescued discs (ESCRT RNAi, UAS-Hh; hhGal4). **n**, Quantification of band density from **m**. HhNp values were normalized to tubulin values in each individual case.



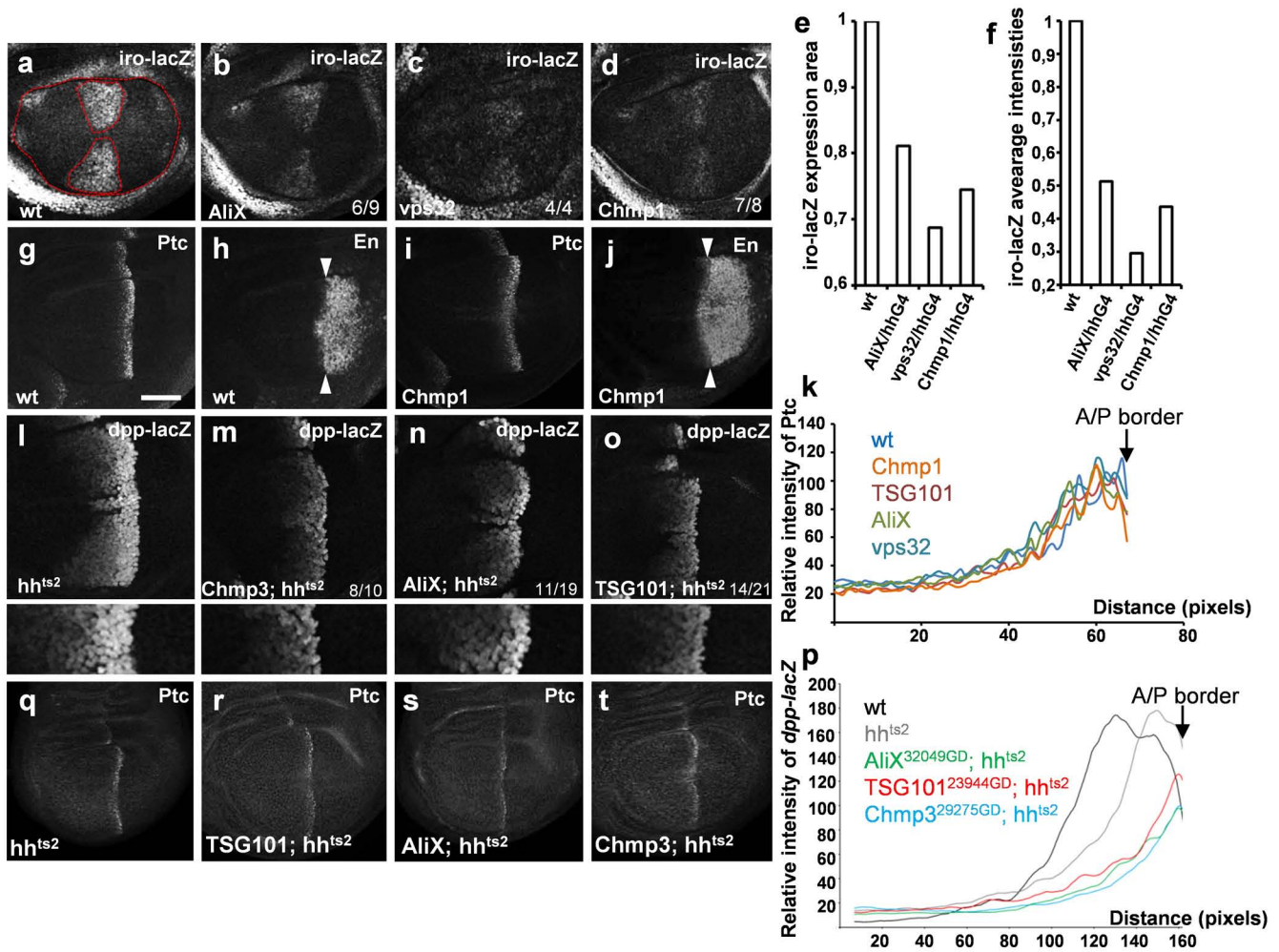
Extended Data Figure 2 | Mild ESCRT RNAi knockdown does not lead to apoptosis or cell polarity changes. **a–g.** Control (**a**, **b**) and ESCRT RNAi (**c–g**) discs stained with antibodies against the apoptotic marker Caspase III (green). Peripodial nuclei are labelled with DAPI (red). Rescued discs with wild-type appearance did not display any sign of apoptosis or cell-architecture defects with the exception of Vps22, investigation of which was not further pursued. Scale bar, 100 μ m. **h–k.** X–Z section of WIDs expressing ESCRT RNAi in the dorsal compartment (D) with apGal4 driver and labelled for the polarity markers DE-cadherin (DCAD2; green) and Disc-large

(Dlg; red). The ventral domain (V) is wild type. Scale bar, 20 μ m. **l–o.** Secretion of Dlp is not affected in discs expressing ESCRT RNAi in the dorsal compartment for 12–30 h. ECDlp, for extracellular Dlp protein. **p–s.** Disp^{HA} subcellular localization is not affected in Hh-producing cells simultaneously expressing ESCRT RNAi for 30 h. In all X–Z images, apical is towards the top and dorsal is to the left. The dorsal–ventral boundary is indicated by a dashed white line. White arrowheads in **p–s** indicate the apical side of the columnar cells.



Extended Data Figure 3 | The anterior expression of Engrailed and Patched is not affected by ESCRT RNAi in conditions of Hh gain of function.
a–l, In all images, dorsal is towards the top and anterior is to the left.

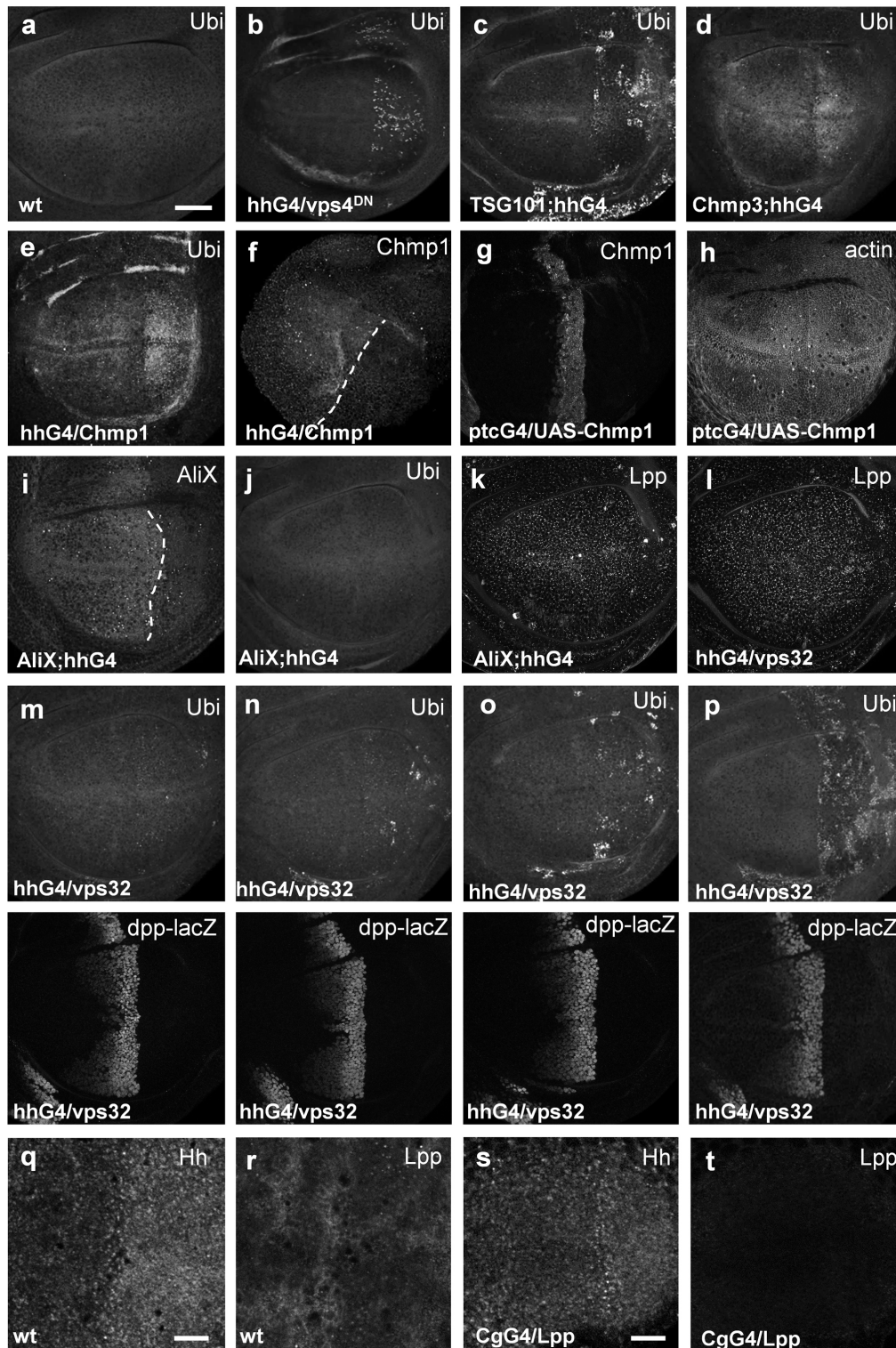
En (a–f) and Ptc (g–l) expression in control discs (a–c, g–i) and discs with tester line plus ESCRT RNAi (d–f, j–l) in the posterior compartment. White triangles indicate the anterior–posterior border. Scale bar, 50 μ m.



Extended Data Figure 4 | Ptc production is maintained in conditions in which *iro-lacZ* and *dpp-lacZ* expression is reduced by ESCRT RNAi.

a–d, The long-range target *iro-lacZ* expression is reduced in discs expressing ESCRT RNAi in the posterior compartment. The red dashed lines indicate the areas used for the quantification in **e**, **f**. **e**, **f**, Quantification of the expression area (ratio of the area of *iro*-expressing cells to the whole WID pouch) (**e**) and intensity (where absolute average intensity values of *iro-lacZ*-expressing cells in the WID pouch were normalized to the background) (**f**) of *iro-lacZ* expression in ESCRT RNAi discs compared with wild type. **g–j**, Ptc (**g**, **i**) and En (**h**, **j**) expression in wild type and conditions of Chmp1 RNAi in posterior cells. White triangles indicate the anterior–posterior (A/P) border. **k**, Quantification of Ptc protein expression in discs expressing ESCRT RNAi in the posterior compartment. **l–o**, Representative examples of the decrease in *dpp-lacZ* expression in *hh^{ts2}* discs with RNAi against ESCRT in posterior cells (*ESCRT RNAi; hh^{ts2}*) grown at restrictive temperature for 4 days from L1

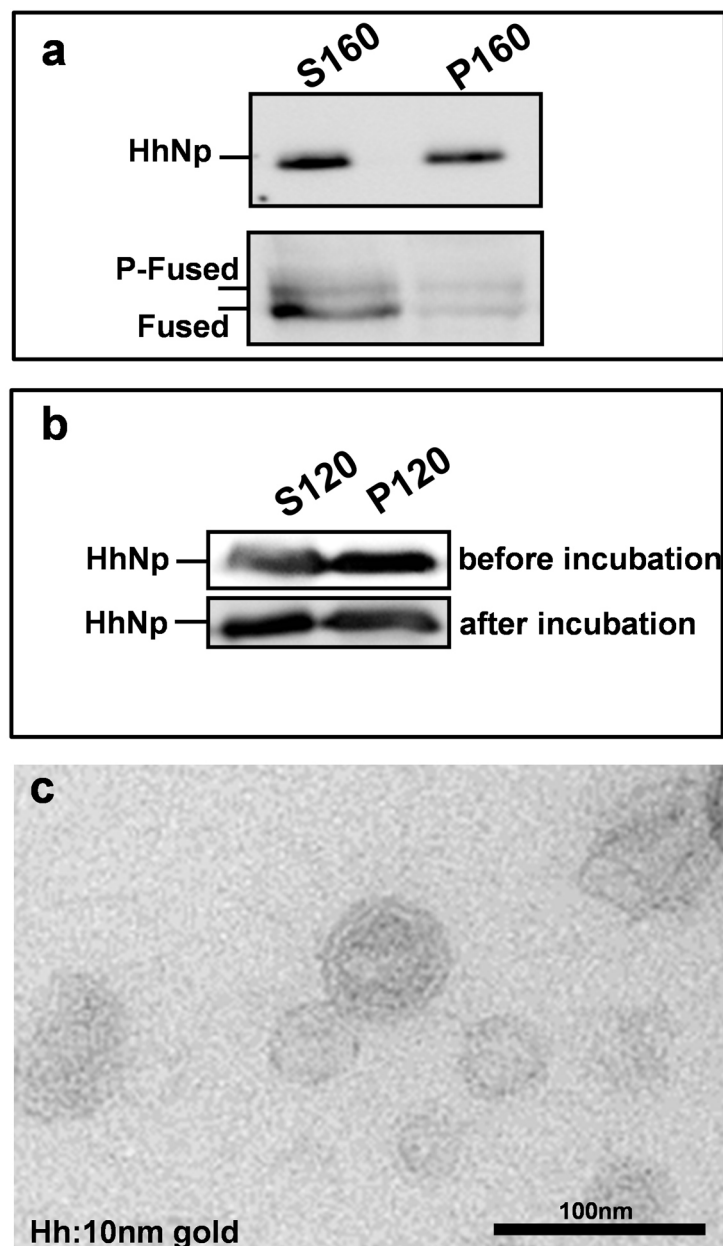
onwards. At restrictive temperature, the *hh^{ts2}* allele is semi-lethal, with up to 60% of the flies dying as pharate adults (Methods). The surviving flies displayed a 50% decrease in the v3–v4 intervein space (Extended Data Fig. 1e, l). *dpp-lacZ* expression, which was mildly affected in the *hh^{ts2}* WID (**l**, **p**), became strongly reduced in all ESCRT RNAi discs in the *hh^{ts2}* background analysed (**m–p**). A magnification of the *dpp-lacZ* stripe is shown for each genotype below the panels. **p**, Quantification of *dpp-lacZ* expression in wild-type, *hh^{ts2}* and *ESCRT RNAi; hh^{ts2}* discs. **q–t**, Ptc expression in conditions of *ESCRT RNAi; hh^{ts2}* discs. ESCRT RNAis had no effect on the decrease of Ptc expression observed in *hh^{ts2}* WIDs, suggesting that Hh long-range activity is sensitive to ESCRT function, but not short-range Hh signalling (**q–t**). On all fluorescence images, dorsal is shown to the top and anterior is to the left. Scale bar, 50 μ m. **b–d**, **l–o**, Numbers on bottom right indicate number of discs displaying reduced *iro* or *dpp* activity per whole population analysed.



Extended Data Figure 5 | Specificity controls for ESCRT RNAi.

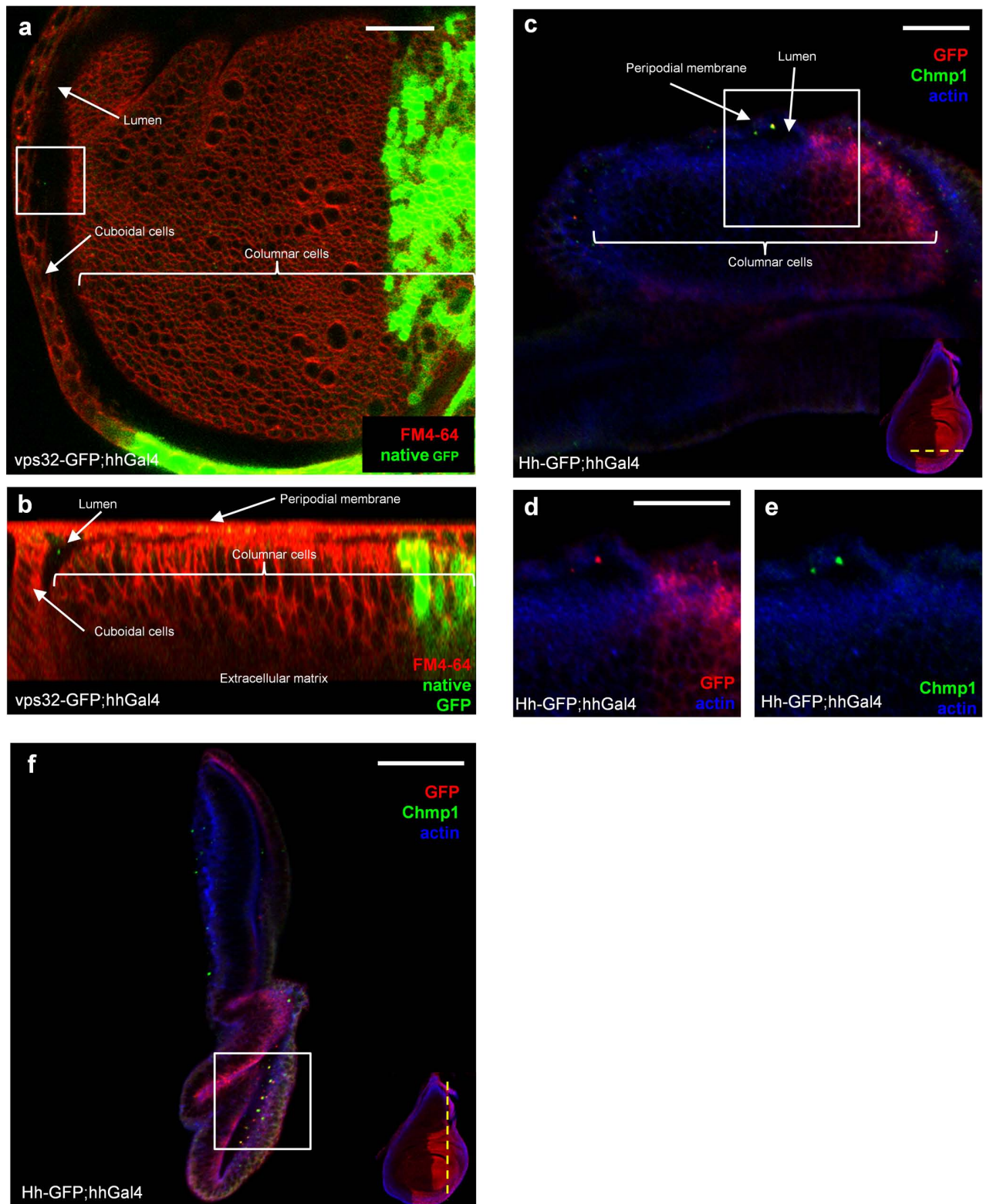
a–e, Expression of *Vps4*^{DN} (**b**) or ESCRT RNAi (**c–e**, and see also in **n–p**) in the posterior compartment results in the accumulation of the poly-Ubi epitope compared with wild type (**a**). **f–h**, *Chmp1* RNAi in the Hh-producing cells leads to the depletion of *Chmp1* epitopes (**f**), whereas overexpression of *UAS-Chmp1* in the *Ptc* domain results in the accumulation of the *Chmp1* signal (**g, h**). **i**, *AliX* RNAi in the Hh-producing cells leads to the depletion of *AliX* epitopes. **j, k**, *AliX* RNAi in Hh-producing cells does not result in the accumulation of either poly-Ubi or *Lpp* epitopes. This disc is the same as the one in Fig. 2b. **l–p**, Panels show four examples of discs expressing RNAi against *Vps32* in the posterior compartment stained for *Lpp* (**l**), poly-Ubi (**m–p**, top row) and *dpp-lacZ* (**m–p**, bottom row). Note that *dpp-lacZ* expression can

be reduced (**m**, bottom panel) even in a disc with no change in poly-Ubi (**m**, top panel) and *Lpp* distribution (**l**). Increased poly-Ubi accumulation is not correlated with a further decrease in *dpp-lacZ* expression. **l, m**, The disc shown is the same as the one in Fig. 2c. **q–t**, Expression of an *Lpp* RNAi construct in the fat body using the *CgGal4* (*CgG4*) driver decreases *Lpp* protein levels in the WID (**r, t**), but does not cause *Hh* accumulation at the apical cell surface (**q, s**). Shown is a close-up of the central disc quadrant overlying the anterior-posterior compartment boundary. In all images dorsal is at the top and anterior is to the left. **s, t**, Note that because of general depletion of *Lpp* the size of the disc is considerably reduced. **a, q, s**, Scale bars, 30 μ m. In all images dorsal is at the top and anterior is to the left.



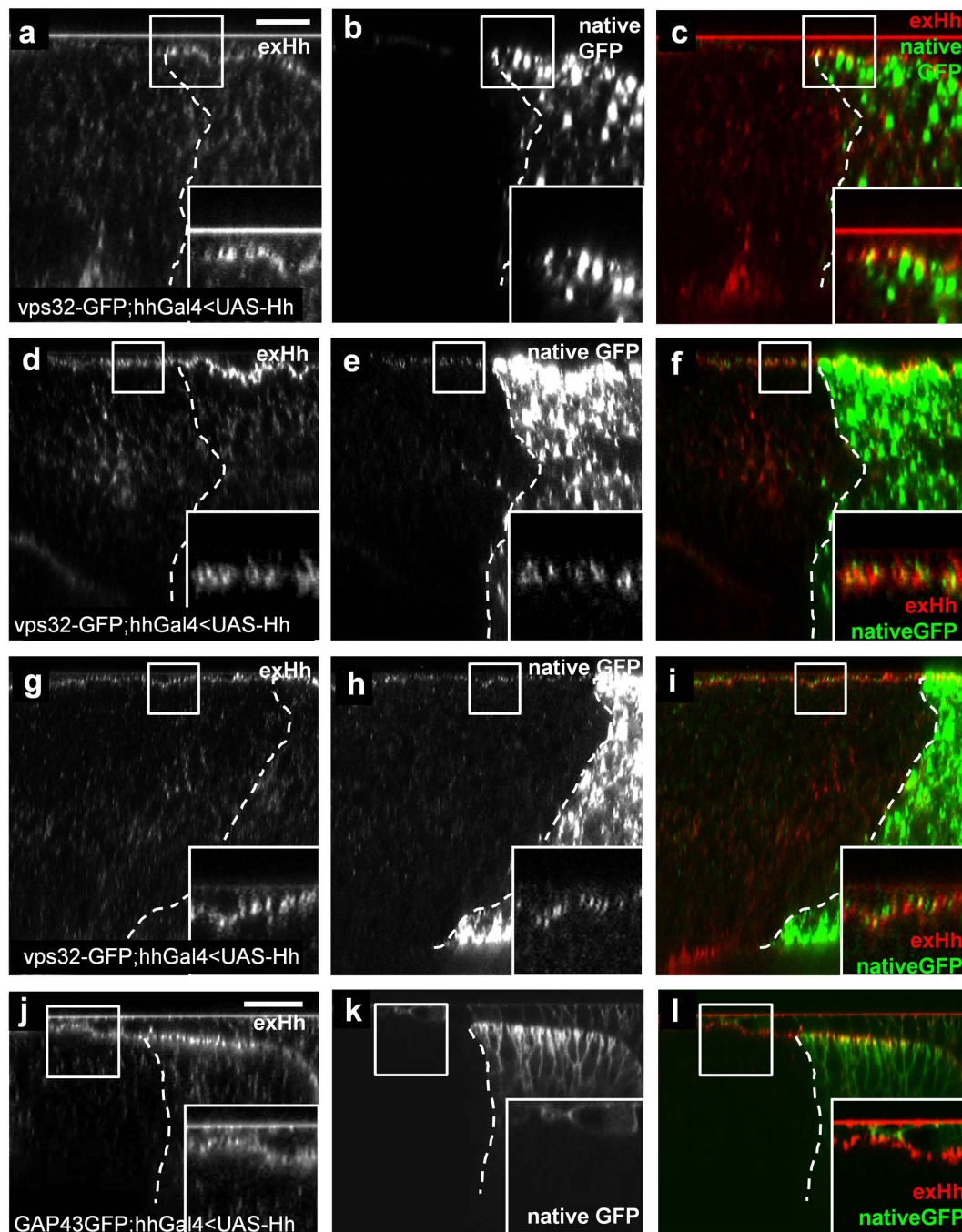
Extended Data Figure 6 | Control experiments for exovesicle characterization. **a**, The S10 fraction was run at 160,000g for 24 h. Hh protein split equally between the 160,000g supernatant and the 160,000g pellet. These two fractions induced Fu phosphorylation (P-Fused) with the same intensity. **b**, The relative amount of S120-Hh and P120-Hh before and after

incubation of cells in Fig. 3c. **c**, Immunogold labelling of P120 fraction isolated from non-Hh-expressing Cl8 cells. In this experimental condition the Hh epitope is not detectable on the surface of exovesicles. This is a control for Fig. 3e.



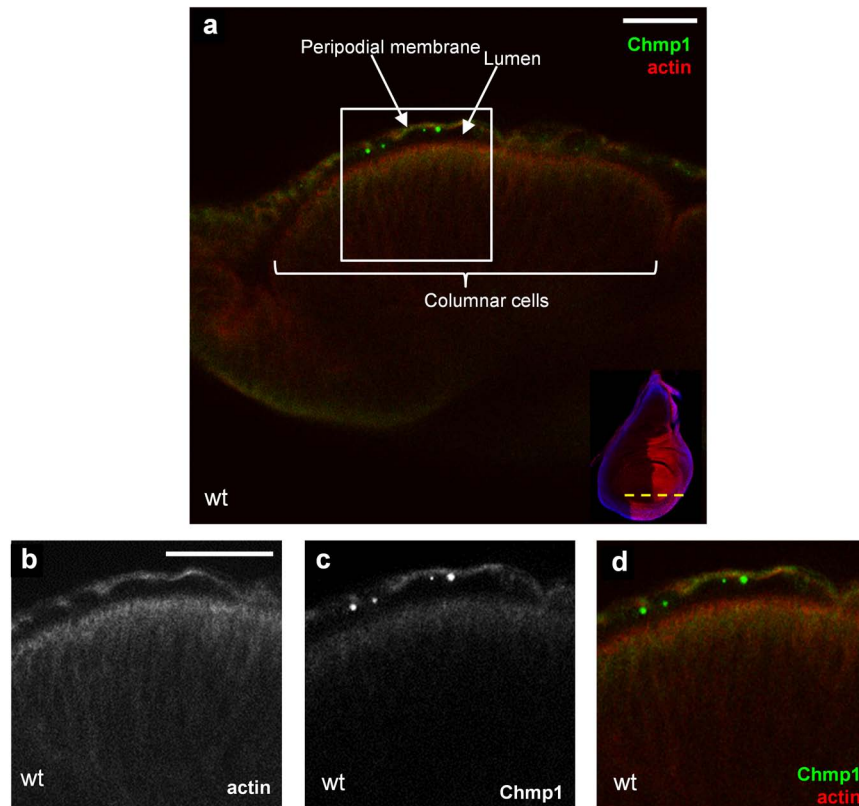
Extended Data Figure 7 | Visualization of ESCRT-positive exovesicles in the luminal space of *Drosophila* WIDs. **a, b,** Live imaging of Vps32-GFP in a WID. Cell membranes are stained with FM4-64 (red). Scale bar, 20 μ m. **a,** X-Y section through the WID pouch showing the posterior Vps32-GFP-producing cells on the right (strong green signal) and the anterior luminal space on the left. The white square indicates the region shown at a higher magnification in Fig. 4b. **b,** X-Z section showing the Vps32-GFP-expressing columnar cells on the right (bright green) and a Vps32-GFP-positive vesicle in the anterior lumen. **c-f,** Colocalization of Hh-GFP and endogenous Chmp1 protein in the WID lumen. **c-e,** Transverse section through the ventral part of the wing pouch

parallel to the dorsal-ventral axis. Scale bars, 20 μ m. **d, e,** Higher magnification views of the region indicated by the white box in **c**. Note the presence of Hh- and Chmp1-positive exovesicles in the luminal space in **c-f**. Cell outlines are visualized by the staining of cortical filamentous actin (phalloidin, blue). **f,** Transverse section through the WID across the dorsal-ventral axis. The white square indicates the region shown at a higher magnification in Fig. 4f-h. Scale bar, 50 μ m. Schematic representations of the positions of the optical sections (dashed yellow lines) used in **c** and **f** are shown in the bottom-right corner of the corresponding images.



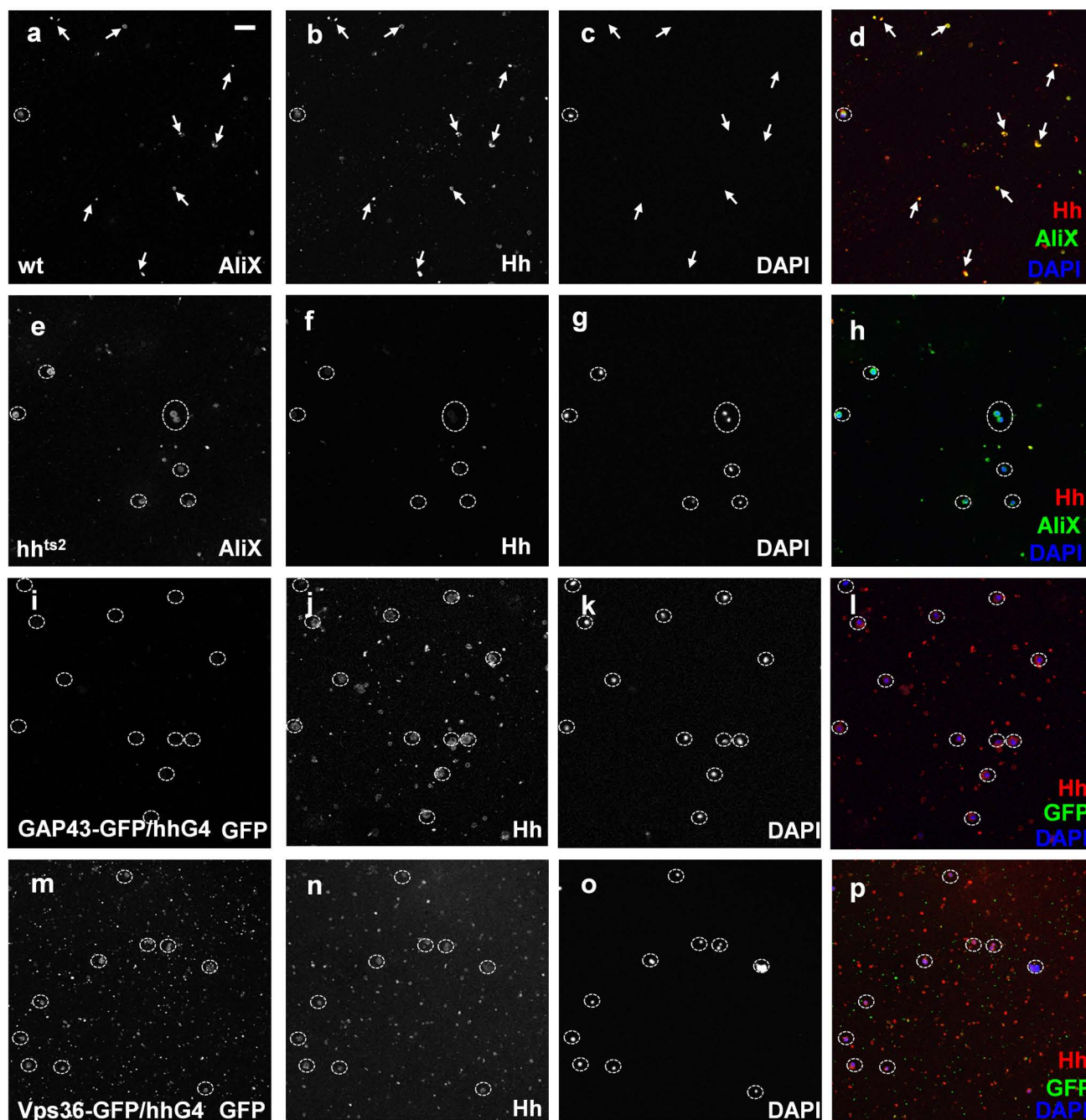
Extended Data Figure 8 | Controls and additional examples for the extracellular colocalization of Hh and Vps32-GFP. a–i, Partial colocalization of extracellular Hh and Vps32-GFP on the apical side of both the producing (a–c) and recipient cells (d–i). Scale bar, 10 μ m. Pearson coefficients: 0.426 (d–f) and 0.542 (g–i). M1M2 coefficients: 0.156 for Hh and 0.146 for Vps32-GFP (d–f); 0.284 for Hh and 0.382 for Vps32-GFP (g–i). j–l, GAP43-GFP, an inner plasma membrane marker, was expressed in the posterior

compartment. No colocalization with Hh was observed in the anterior compartment (Pearson coefficient: 0.009; M1M2 coefficients: 0.001 for Hh and 0.142 for GAP43-GFP). Note that GAP43-GFP-labelled peripodial membrane above the columnar cells. Scale bar, 20 μ m. Apical is towards the top and anterior is to the left. The anterior–posterior border is marked with a dashed white line. Magnification of Hh punctae is shown in the bottom-right corner.



Extended Data Figure 9 | Endogenous Chmp1 is secreted into the luminal space of WID. **a–d**, Transverse section through the ventral part of the wing pouch across the anterior–posterior axis. Scale bars, 20 μm . **b–d**, Higher magnification views of the region indicated by the white box in **a**. Note the

presence of Chmp1-positive exovesicles in the luminal space. Cell outlines are visualized by the staining of cortical filamentous actin (phalloidin, red). Schematic representation of the position of the optical section (dashed yellow line) used in **a–d** is shown in the bottom-right corner of **a**.



Extended Data Figure 10 | Specificity controls for *Drosophila* haemolymph staining. **a–d**, Endogenous Hh colocalizes with AliX on non-cellular (DAPI-negative) particles in the *Drosophila* haemolymph. White arrows indicate examples of colocalization. **e–h**, Most of the Hh epitope is depleted from the circulating particles in the *hh^{ts2}* homozygous mutant, which was kept

at a restrictive temperature for 4 days. **i–l**, GAP43–GFP expressed in all posterior cells is not secreted into the *Drosophila* circulation. **m–p**, The ESCRT II member Vps36–GFP expressed in all posterior cells does not colocalize with endogenous Hh in the fly blood, but is secreted as different particles instead. Scale bar, 20 μ m. White circles mark DAPI-positive blood cells.

Centriole amplification by mother and daughter centrioles differs in multiciliated cells

Adel Al Jord^{1,2,3}, Anne-Iris Lemaître^{1,2,3}, Nathalie Delgehyr^{1,2,3}, Marion Faucourt^{1,2,3}, Nathalie Spassky^{1,2,3*} & Alice Meunier^{1,2,3*}

The semi-conservative centrosome duplication in cycling cells gives rise to a centrosome composed of a mother and a newly formed daughter centriole¹. Both centrioles are regarded as equivalent in their ability to form new centrioles and their symmetric duplication is crucial for cell division homeostasis^{2–4}. Multiciliated cells do not use the archetypal duplication program and instead form more than a hundred centrioles that are required for the growth of motile cilia and the efficient propelling of physiological fluids⁵. The majority of these new centrioles are thought to appear *de novo*, that is, independently from the centrosome, around electron-dense structures called deuterosomes^{6–8}. Their origin remains unknown. Using live imaging combined with correlative super-resolution light and electron microscopy, we show that all new centrioles derive from the pre-existing progenitor cell centrosome through multiple rounds of procentriole seeding. Moreover, we establish that only the daughter centrosomal centriole contributes to deuterosome formation, and thus to over ninety per cent of the final centriole population. This unexpected centriolar asymmetry grants new perspectives when studying cilia-related diseases^{5,9} and pathological centriole amplification observed in cycling cells and associated with microcephaly and cancer^{2–4,10}.

Large-scale centriole production in multiciliated cells is thought to occur through two independent pathways: a commonly described ‘*de novo*’ pathway, also known as the acentriolar/deuterosome pathway; and a centriolar pathway, where new centrioles are formed adjacent to pre-existing ones in a manner akin to centriole duplication^{6,11,12}. The significance of having two pathways to produce the same organelle is enigmatic and nothing is known about their interplay or their exact contribution to the final population of centrioles. We first established that both pathways occur in the brain ependyma as in other mammalian ciliated epithelia^{7,11–13} (Fig. 1a, b and Extended Data Fig. 1a). Using live imaging, we traced the dynamics of centriole formation in ependymal cells from transgenic mice expressing a GFP-tagged version of the distal core centriolar protein centrin 2 (ref. 14), Cen2-GFP (Cen2 also known as Cctn2). At the onset of centriole amplification, a Cen2-GFP cloud accumulated around the pre-existing centrosome (Fig. 1c, d), which spawned a growing number of Cen2-GFP rings, or ‘halos’, that accumulated in the cytoplasm (Fig. 1e, f). The Cen2-GFP fluorescence in every halo intensified progressively and adopted a flower-like shape (a ring of fluorescence resembling petals; Fig. 1g). The flowers finally dissociated into tens of Cen2-GFP dots (Fig. 1h). The Cen2-GFP⁺ centrioles then migrated to the apical membrane where they initiated the extension of the motile ciliary tufts (Extended Data Fig. 2a, b and Supplementary Videos 1 and 2; $n = 12$). Together with immunostainings of centriolar markers on fixed developing ependymal cells (Extended Data Fig. 1b–h), these observations show that centriole amplification occurs in the vicinity of the centrosome and that the ‘halo stage’, which corresponds to the accumulation of centriolar precursors, is followed by a ‘flower stage’, during which maturing centrioles appear in a synchronized manner.

To characterize the origin and composition of the halos further, we monitored centrosomal centrioles at a higher resolution. Live imaging showed that nascent halos budded from the wall of one centrosomal

centriole in the form of Cen2-GFP⁺ rings that detached and accumulated in the nearby cytoplasm (Fig. 2a, Extended Data Fig. 2c, d and Supplementary Videos 3 and 4; $n = 15$ cells). Use of 3D-structured illumination microscopy (3D-SIM) showed that the halos were composed of Sas-6⁺ (also known as Sass6)/Cen2-GFP⁺ subunits (Fig. 2b, c, Extended Data Fig. 2e and Supplementary Video 5) indicating that they represent spherical assemblies of procentrioles. A generational asymmetry exists between centrosomal centrioles and has been shown to confer different functions to each member of the pair^{15–17}. To determine the contribution of mother and daughter centrioles to procentriole formation, we immunostained the structural protein Sas-6 (ref. 18) and the key regulators of centriole biogenesis Cep152 and Plk4 (refs 19,20). In ependymal cycling progenitors (positive for the cycling cell marker Ki67) duplicating their centrosome, these proteins stained symmetrically on both centrosomal centrioles. However, in postmitotic differentiating cells (positive for the multiciliated cell transcription factor Foxj1) the immunostains were significantly larger and more intense on the daughter than on the mother centriole (Fig. 2d, e and Extended Data Fig. 2f, g). Additionally, conventional and 3D-SIM analyses revealed that 100% of the Cen2-GFP⁺ halos formed on the wall of the daughter centriole ($n = 33$) and were contiguous to the proximal portion of the centriole (Fig. 2f–h), which is the canonical site of procentriole nucleation in cycling cells. Although the onset of differentiation was characterized by the lengthening of the primary cilium (a cell antennae nucleated by the mother centriole²¹), no alteration of centrosome asymmetry was observed in ciliary mutants throughout the amplification process (Extended Data Fig. 9a–d). The centrosome thus switches from symmetric canonical centriole duplication in cycling progenitors to asymmetric centriole amplification in

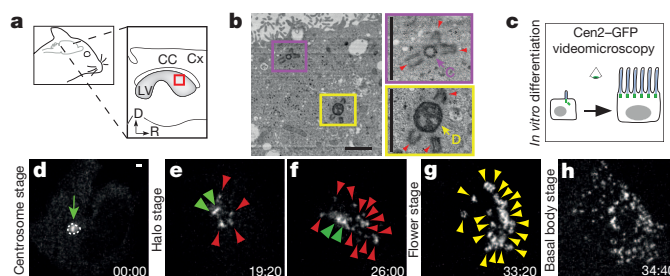


Figure 1 | Centriole amplification dynamics during ependymal cell differentiation. **a**, Sagittal view of the mouse lateral ventricular wall. Square indicates the region analysed in **b**. Cx, cortex; CC, corpus callosum; LV, lateral ventricle; D/R, dorsal/rostral. **b**, Transmission electron microscopy (TEM) on a differentiating P4 ependymal progenitor cell showing procentrioles (red arrowheads) growing from a centriole (C, ‘centriolar’ pathway) or from a deuterosome (D, ‘deuterosome’ pathway). **c–h**, Videomicroscopy during Cen2-GFP⁺ ependymal progenitor differentiation. **d–h**, From Supplementary Video 1. Arrow indicates the centrosome; dashed line delineates the Cen2-GFP⁺ cloud; green arrowheads point to centrosomal centrioles; red arrowheads point to halos; yellow arrowheads point to flowers. $\bar{t}_{\text{amplification}} = 29.3 \pm 6$ h, $\bar{t}_{\text{halos}} = 21 \pm 5$ h and $\bar{t}_{\text{flowers}} = 9.7 \pm 2$ h, where \bar{t} = mean time. Time in hh:mm; scale bars, 1 μ m.

¹Ecole Normale Supérieure, Institut de Biologie de l’ENS, IBENS, F-75005 Paris, France. ²Inserm, U1024, F-75005 Paris, France. ³CNRS, UMR 8197, F-75005 Paris, France.

*These authors contributed equally to this work.

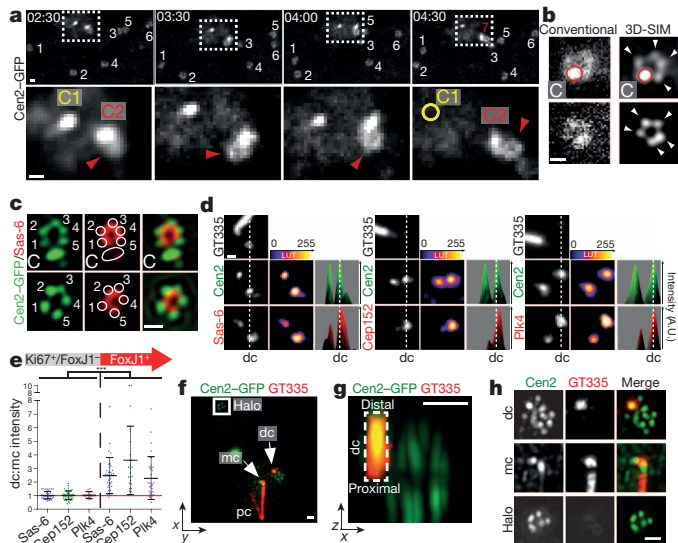


Figure 2 | Halos are a spherical assembly of procentrioles nucleated from the centrosomal daughter centriole. **a**, From Supplementary Video 3. Cen2-GFP dynamics during halo formation. Top: z-projections (4.2 μm) of centrosomal centrioles (C1 and C2) and six halos in the cytoplasm. Bottom: single z-plane (0.7 μm) of the centrosome. Arrowhead points to the seventh halo forming from C2. **b**, Single 0.1 μm z-plane of a Cen2-GFP⁺ centrosomal (C; top) and cytoplasmic halo (bottom) imaged by conventional epifluorescence microscopy (left) and 3D-SIM (right). Arrowheads indicate Cen2-GFP⁺ subunits observed only with 3D-SIM. **c**, 3D-SIM z-plane of a Cen2-GFP⁺ centrosomal (C; top) or cytoplasmic (bottom) halo immunostained with Sas-6 (red). Circles delineate Cen2-GFP⁺ subunits. **d**, Sas-6, Cep152 and Plk4 stainings on Cen2-GFP⁺ mother (growing GT335⁺ primary cilium) and daughter (dc) centrosomal centrioles at the halo stage. 'Fire' colour look-up table (LUT) and fluorescence intensity profiles are shown. A.U., arbitrary units. **e**, Symmetric to asymmetric switch of centrosomal markers. Daughter:mother centriole (dc:mc) signal ratios in cycling Ki67⁺/Foxj1⁺ (left; Sas-6 ($n = 42$), Cep152 ($n = 43$), Plk4 ($n = 18$)) and differentiating Foxj1⁺ ependymal progenitors at the halo stage (right; Sas-6 ($n = 33$), Cep152 ($n = 25$), Plk4 ($n = 32$)). Error bars represent mean \pm s.d.; P values derived from two-tailed Mann-Whitney U -tests, *** $P \leq 0.0003$; red line represents a ratio equal to 1. **f**, 3D-SIM z-projection (1.8 μm) of Cen2-GFP and GT335 immunoreactivity at the halo stage. Centrin aggregates are occasionally observed in Cen2-GFP cells. **g**, Side view of the GT335⁺ daughter centriole from **f** showing a Cen2-GFP⁺ halo contiguous to the proximal portion of the daughter centriole. **h**, z-projection (0.6 μm) magnifications of both centrosomal centrioles (top and centre) and the cytoplasmic halo (bottom) from **f**. pc, primary cilium; time in hh:mm; scale bars, 0.5 μm .

postmitotic ependymal progenitors. The amplification is a recurrent process during which procentrioles are nucleated from the centrosomal daughter centriole to form halos that are released into the cytoplasm.

Deuterosomes are the source of the presumed 'de novo' pathway⁶, and are composed of two newly identified proteins, Deup1 and Ccdc78, which recruit Cep152 and Plk4 to drive centriole amplification^{7,8}. Deup1 and Ccdc78 were localized to the centrosome in ependymal cycling progenitors (Extended Data Fig. 3a). As the cells entered the halo stage, substantial accumulation of both proteins was observed at the daughter centriole (Extended Data Fig. 3b, c). Since high concentrations of Deup1 were shown to trigger deuterosome formation⁷, amplification from the daughter centriole may be driven by the formation of deuterosomes. Consistently, 3D-SIM and live imaging highlighted the presence of a Deup1⁺ cluster in the centre of the halo released by the daughter centriole (Fig. 3a, b, Extended Data Fig. 3d and Supplementary Video 6). To test this hypothesis further, we correlated 3D-SIM observations with transmission electron microscopy on Cen2-GFP⁺ cells (3D-SIM/EM; Extended Data Fig. 3e). 3D-SIM/EM on halo-stage cells showed that the Cen2-GFP⁺ subunits of the halos were actually procentrioles organized around spherical deuterosomes (Fig. 3c). Serial sections of entire cells ($n = 3$ cells)

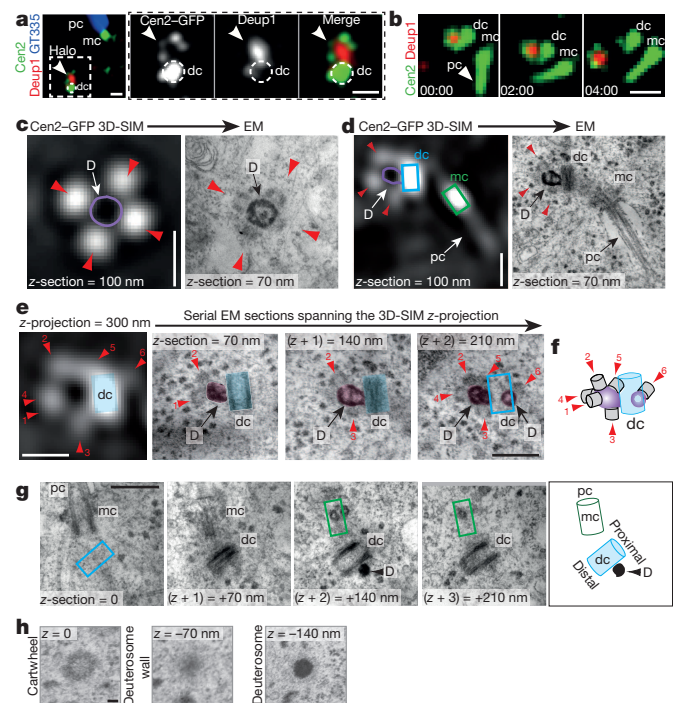


Figure 3 | Immature procentrioles are amplified from the daughter centriole via deuterosome formation. **a**, 3D-SIM z-plane (0.12 μm) of a Cen2-GFP⁺ centrosome during halo (arrowhead) formation from the daughter centriole (circle) immunostained with Deup1 and GT335. Dotted panel shows a z-projection (0.6 μm) magnification of the daughter centriole. dc, daughter centriole; mc, mother centriole; pc: primary cilium. **b**, Cen2-GFP/ Kusabira-Orange-Deup1 time-lapse images during halo release from the daughter centriole. Note that Cen2-GFP occasionally enters the primary cilium. Time in hh:mm. **c**, **d**, Correlative images of 3D-SIM z-plane (left) and TEM (right) of a cytoplasmic halo (c) showing procentrioles on a deuterosome, and of a centrosome (d) showing an incomplete procentriole-loaded deuterosome connected to the daughter centriole. Purple lines delineate TEM deuterosomes. D, deuterosome. **e**, Correlative images of 3D-SIM z-plane of the daughter centriole (blue) from **d** and the corresponding serial EM sections spanning the deuterosome (purple). Rectangle indicates the daughter centriole location on the previous EM section. A second, smaller deuterosome is connected to another site on the proximal wall of the daughter centriole. Spatial resolution of 3D-SIM is insufficient to individualize some of the procentrioles visible by EM. **f**, 3D-reconstruction of **e**. **g**, Serial EM sections of the centrosome in an early differentiating ependymal progenitor showing the first deuterosome without procentrioles adjacent to the daughter centriole. The image on the right shows an illustration of the collated EM images. **h**, Serial EM sections of a procentriole at the halo stage growing from a deuterosome. Red arrowheads, procentrioles; scale bars, 50 nm (**h**) and 0.5 μm (**a–g**).

showed that all Cen2-GFP⁺ procentrioles were associated with deuterosomes and, conversely, that all deuterosomes were loaded with procentrioles (23 out of 23) (Extended Data Figs 4 and 5). Developing halos corresponded to deuterosomes with incomplete spherical shapes attached to the proximal segment of the daughter centriole (Fig. 3d and Extended Data Figs 4 and 5). One or two deuterosomes were observed connected to the daughter centriole (Fig. 3e, f and Extended Data Figs 4, 5 and 6b, c). To visualize the first events of deuterosome formation, we screened early ependymal progenitors, identified by the absence of deuterosomes in the cytoplasm and a small accumulation of electron-dense aggregates⁶, and found cells where the daughter centriole had a single attached deuterosome without detectable procentrioles (Fig. 3g and Extended Data Fig. 6a; $n = 3$ cells). The procentriole-free deuterosomes were smaller than centrosome-bound deuterosomes loaded with procentrioles and, more generally, deuterosome size was correlated with the number of procentrioles they support (Extended Data Fig. 6d, e). Deuterosomes are

thus generated from the proximal segment of the daughter centriole and do not appear *de novo*.

During the halo stage/deuterosome formation, procentrioles displayed a tube-like structure organized around a cartwheel containing microtubule singlets (Fig. 3h). The absence of microtubule triplets indicated that the procentrioles were immature. Consistently, the lengths and widths of procentrioles were significantly smaller than those of mature centrioles. Procentrioles were similar in size both within and among halo-stage cells (Fig. 4g, h, red and grey quantifications; Extended Data Figs 7 and 8e) and were comparable to early S-phase procentrioles in cycling cells²².

Together, these data show that during ependymal cell development the centrosomal daughter centriole greatly amplifies procentrioles by generating intermediate structures, the deuterosomes. Once nucleated at the daughter centriole, procentrioles remain latent at an early stage of their biogenesis on the deuterosomes.

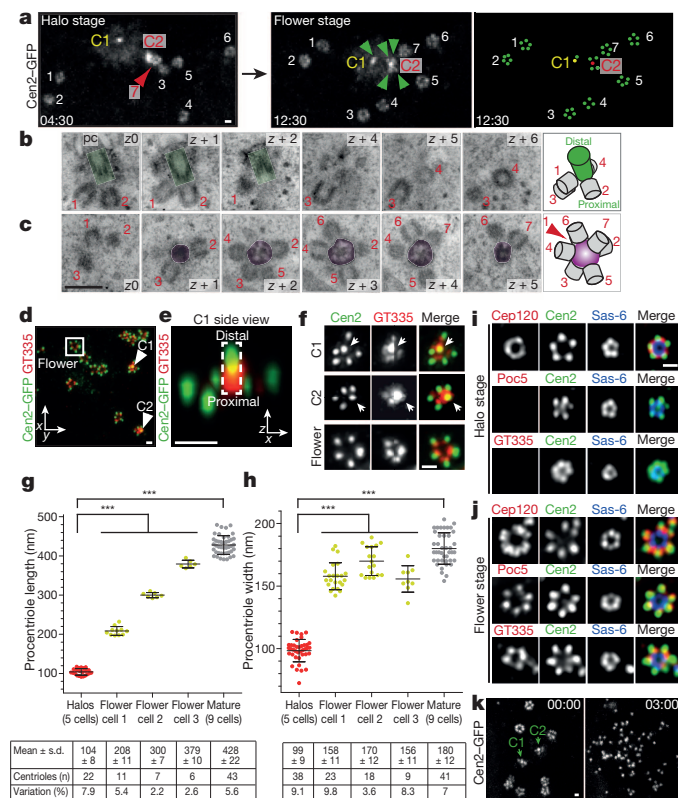


Figure 4 | Procentrioles grow simultaneously from both deuterosomes and centrosomal centrioles after formation of the last deuterosome. **a**, From Supplementary Video 3. Cen2-GFP dynamics showing the transformation of halos (numbered at 04:30) into flower (numbered at 12:30). Red arrowhead indicates the last halo (seventh) to be formed (same cell as Fig. 2a). Green arrowheads point to procentrioles on the wall of centrosomal centrioles. Right panel illustrates procentrioles (green) and C1/C2 centrosomal centrioles. **b, c**, Serial EM sections spanning a centrosomal centriole (**b**; green) and a deuterosome (**c**; purple) in a flower-stage cell. Procentrioles are numbered. Images on the right show 3D-reconstruction schematics. **d**, 3D-SIM z-projection (1.8 μ m) of Cen2-GFP and GT335 immunoreactivity at the flower stage. **e**, Side view of C1 showing procentrioles growing from its proximal portion. **f**, z-projection (0.6 μ m) magnifications of C1/C2 centrosomal centrioles (white arrows) and deuterosome (flower). **g, h**, Lengths and widths of procentrioles from cells at the halo, flower and mature basal body stages; error bars represent mean \pm s.d.; *P* values derived from two-tailed Mann-Whitney *U*-tests, ****P* < 0.0001. **i, j**, 3D-SIM z-planes (120 nm) comparing Cep120, Poc5 and GT335 immunoreactivities between halos (**i**) and flowers (**j**). **k**, From Supplementary Video 7. Cen2-GFP dynamics showing simultaneous release of procentrioles from centrosomal centrioles and deuterosomes. Time in hh:mm; scale bars, 0.5 μ m.

We next visualized the dynamics of centriole growth during the halo-to-flower stage transition. Immediately after the formation of the last halo/deuterosome at the daughter centriole, all of the Cen2-GFP⁺ halos simultaneously transformed into intensely fluorescent flower-like structures. At this stage, up to five procentrioles were also seen growing directly from the proximal portions of both centrosomal centrioles (Fig. 4a–f, Extended Data Figs 6, 8a, b, d and Supplementary Videos 3 and 4; *n* = 30 cells). These ‘centriolar’ procentrioles were probably nucleated during the halo stage, since they were already visible during the late phases of deuterosome formation (Fig. 2h and Extended Data Figs 4, 5 and 6b). Both centrosomal centrioles thus produce procentrioles with a canonical orthogonal configuration, which defines the so-called ‘centriolar’ pathway. However, more than 90% of the centrioles were generated via deuterosomes and less than 10% directly from centrosomal centrioles (14 ± 1 deuterosomes per cell, 4.2 ± 0.4 procentrioles per centrosomal centriole, 5.6 ± 0.3 per deuterosome (mean \pm s.e.m.); *n* = 10 cells).

All procentrioles in a given cell grew synchronously during the flower stage, since their sizes were similar in each cell and increased significantly in comparison to the procentrioles in halo-stage cells (Fig. 4g, h and Extended Data Figs 6f, 7 and 8d, e). At the molecular level, halo-stage procentrioles were positive for the early markers of centriole biogenesis (Cep152, Plk4, Sas-6, Stil, Ckap (also known as Cenpj), Cep120, CP110), and flower procentrioles became positive for markers of the late steps of centriole assembly (Poc5 (ref. 23) and GT335 (ref. 24) (an antibody against glutamylated tubulin)) (Fig. 4i, j and Extended Data Fig. 8f, g). Centriole biogenesis in multiciliated cells is therefore a two-step process in which procentriole initiation (halo stage/deuterosome formation) is clearly demarcated from the late phases of centriole assembly (flower stage). Synchronized growth is followed by the simultaneous detachment of procentrioles from both centrosome and deuterosome platforms ($\bar{t} = 2.5 \pm 0.85$ h; *n* = 9 cells; Fig. 4k, Extended Data Fig. 8c and Supplementary Videos 4 and 7). After migration and docking at the apical membrane, Sas-6 disappears from centrioles as motile cilia grow (Extended Data Fig. 8h). Identical steps of centriole amplification and similar centrosome asymmetry were observed in tracheal multiciliated cells *in vivo* (Extended Data Fig. 9e–k), suggesting that all mammalian multiciliated cells might share this mechanism of large-scale centriole biogenesis.

Our results revise the ‘*de novo*’ hypothesis; the so-called ‘centriolar’ and ‘deuterosome’ pathways are the outcomes of a single pathway orchestrated by the centrosome (Fig. 5). The ability of mother and daughter centrioles to differentially control both the number of and the way they generate new centrioles reveals a new centrosome asymmetry. A differential regulation of Deup1 levels at centrosomal centrioles could underlie the ability of the daughter centriole to form deuterosomes and amplify centrioles⁷. The identification of proteins that interact with Deup1 using proximity-dependent biotinylation²⁵ may help uncover the upstream signals controlling centrosome asymmetry in these cells.

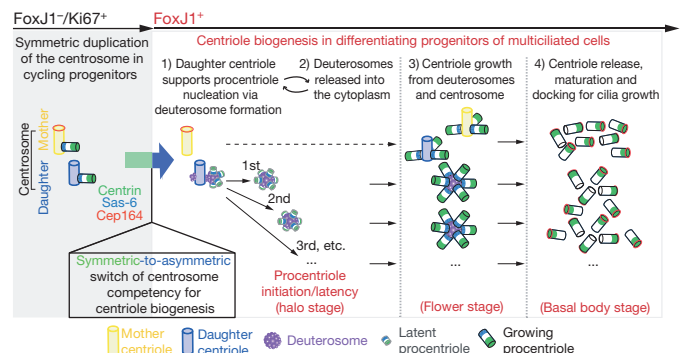


Figure 5 | Model of centriole amplification in multiciliated cells.

Deuterosomes develop from the same proximal region of the centriolar wall from which procentrioles grow during canonical centriole duplication and use known regulators of this process^{7,8} (Extended Data Fig. 10). They thus function as shuttles, displacing the late phases of centriole assembly from the centrosome to the cytoplasm while procentriole maturation is arrested. This mechanism allows massive, albeit centriole-mediated, production of centrioles. Emerging studies showing the involvement of S-phase regulators during multiciliated cell differentiation^{26–30}, together with the phase-like progression of centriole amplification (Fig. 5), suggest that the amplification process may be a facet of the duplication cycle, in which, however, spatial and temporal controls of centriole number are bypassed. It would therefore be interesting to explore the roles of S, G2 and mitosis regulators throughout the amplification stages delimited here.

In this study, we report that, under physiological conditions, the centrosome supports either the duplication of centrioles or their massive amplification. Our finding unifies the mechanisms of centriole biogenesis in cycling and multiciliated cells. Since the daughter centriole of the centrosome is the primary nucleation point for amplification, it could play a major role in the control of centriole production. This newly found centrosome asymmetry thus provides a new paradigm for understanding the regulation of centriole biogenesis and number.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 January; accepted 14 August 2014.

Published online 12 October 2014.

- Vorobjev, I. A. & Chentsov, Yu. S. Centrioles in the cell cycle. I. Epithelial cells. *J. Cell Biol.* **93**, 938–949 (1982).
- Lingle, W. L. *et al.* Centrosome amplification drives chromosomal instability in breast tumor development. *Proc. Natl Acad. Sci. USA* **99**, 1978–1983 (2002).
- Ganem, N. J., Godinho, S. A. & Pellman, D. A mechanism linking extra centrosomes to chromosomal instability. *Nature* **460**, 278–282 (2009).
- Marthiens, V. *et al.* Centrosome amplification causes microcephaly. *Nature Cell Biol.* **15**, 731–740 (2013).
- Sharma, N., Berbari, N. F. & Yoder, B. K. Ciliary dysfunction in developmental abnormalities and diseases. *Curr. Top. Dev. Biol.* **85**, 371–427 (2008).
- Sorokin, S. P. Reconstructions of centriole formation and ciliogenesis in mammalian lungs. *J. Cell Sci.* **3**, 207–230 (1968).
- Zhao, H. *et al.* The Cep63 paralogue Deup1 enables massive *de novo* centriole biogenesis for vertebrate multiciliogenesis. *Nature Cell Biol.* **15**, 1434–1444 (2013).
- Klos Dehning, D. A. *et al.* Deuterosome-mediated centriole biogenesis. *Dev. Cell* **27**, 103–112 (2013).
- Wallmeier, J. *et al.* Mutations in CCNO result in congenital mucociliary clearance disorder with reduced generation of multiple motile cilia. *Nature Genet.* **46**, 646–651 (2014).
- Godinho, S. A. *et al.* Oncogene-like induction of cellular invasion from centrosome amplification. *Nature* **510**, 167–171 (2014).
- Anderson, R. G. & Brenner, R. M. The formation of basal bodies (centrioles) in the rhesus monkey oviduct. *J. Cell Biol.* **50**, 10–34 (1971).
- Dirksen, E. R. Centriole morphogenesis in developing ciliated epithelium of the mouse oviduct. *J. Cell Biol.* **51**, 286–302 (1971).
- Loots, G. P. & Nel, P. P. Early stages of ciliogenesis in the respiratory epithelium of the nasal cavity of rabbit embryos. *Cell Tissue Res.* **255**, 589–594 (1989).
- Higginbotham, H., Bielas, S., Tanaka, T. & Gleeson, J. G. Transgenic mouse line with green-fluorescent protein-labeled Centrin 2 allows visualization of the centrosome in living cells. *Transgenic Res.* **13**, 155–164 (2004).
- Piel, M., Meyer, P., Khodjakov, A., Rieder, C. L. & Bornens, M. The respective contributions of the mother and daughter centrioles to centrosome activity and behavior in vertebrate cells. *J. Cell Biol.* **149**, 317–330 (2000).
- Wang, X. *et al.* Asymmetric centrosome inheritance maintains neural progenitors in neocortex. *Nature* **461**, 947–955 (2009).
- Anderson, C. T. & Stearns, T. Centriole age underlies asynchronous primary cilium growth in mammalian cells. *Curr. Biol.* **19**, 1498–1502 (2009).
- Kitagawa, D. *et al.* Structural basis of the 9-fold symmetry of centrioles. *Cell* **144**, 364–375 (2011).
- Dzhindzhev, N. S. *et al.* Asterless is a scaffold for the onset of centriole assembly. *Nature* **467**, 714–718 (2010).
- Habedanck, R., Stierhof, Y.-D., Wilkinson, C. J. & Nigg, E. A. The Polo kinase Plk4 functions in centriole duplication. *Nature Cell Biol.* **7**, 1140–1146 (2005).
- Singla, V. & Reiter, J. F. The primary cilium as the cell's antenna: signaling at a sensory organelle. *Science* **313**, 629–633 (2006).
- Robbins, E., Jentzsch, G. & Micali, A. The centriole cycle in synchronized HeLa cells. *J. Cell Biol.* **36**, 329–339 (1968).
- Azimzadeh, J. *et al.* hPOC5 is a centrin-binding protein required for assembly of full-length centrioles. *J. Cell Biol.* **185**, 101–114 (2009).
- Wolff, A. *et al.* Distribution of glutamylated alpha and beta-tubulin in mouse tissues using a specific monoclonal antibody, GT335. *Eur. J. Cell Biol.* **59**, 425–432 (1992).
- Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196**, 801–810 (2012).
- Hoh, R. A., Stowe, T. R., Turk, E. & Stearns, T. Transcriptional program of ciliated epithelial cells reveals new cilium and centrosome components and links to human disease. *PLoS ONE* **7**, e2166 (2012).
- Stubbs, J. L., Vladar, E. K., Axelrod, J. D. & Kintner, C. Multicilin promotes centriole assembly and ciliogenesis during multiciliate cell differentiation. *Nature Cell Biol.* **14**, 140–147 (2012).
- Tan, F. E. *et al.* Myb promotes centriole amplification and later steps of the multiciliogenesis program. *Development* **140**, 4277–4286 (2013).
- Marcet, B. *et al.* Control of vertebrate multiciliogenesis by miR-449 through direct repression of the Delta/Notch pathway. *Nature Cell Biol.* **13**, 693–799 (2011).
- Ma, L., Quigley, I., Omran, H. & Kintner, C. Multicilin drives centriole biogenesis via E2f proteins. *Genes Dev.* **28**, 1461–1471 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Rostaing of the IBENS Electron microscopy facility for making the correlative 3D-SIM and electron microscopy experiments possible. We also thank the Imaging Facility of IBENS, which is supported by grants from Fédération pour la Recherche sur le Cerveau, Région Ile de France DIM NeRF 2009 and 2011 and France-BioImaging. We wish to thank L. Sengmanivong from the Nikon Imaging Centre at Institut Curie-CNRS. We are grateful to L. Goldstein for providing us with the Kif3a mutant mice. We thank C. Janke for the GT335 antibody, E. A. Nigg for the Cep164 and Plk4 antibodies, T. K. Tang for Cep120 and Cpap antibodies, X. Yan and X. Zhu for Deup1 and Cep152 antibodies, J. Azimzadeh for the Poc5 antibody and X. Morin for his gift of the pCAAGS-Cen2-GFP-mCherry plasmid. We also thank M. Bornens, B. Guirao, J. B. Boulé, C. Janke and J.-F. Brunet for their comments on the manuscript and all the members of the Spassky laboratory for discussions. This study was supported by the CNRS, the ENS, INSERM, an FPGG grant, ANR award ANR-12-BSV4-0006 CILIASTEM, ARC award PJA-20131200184, La ligue contre le cancer-comité de Paris RS14/75-88, the 'Investissements d'Avenir' program of the French Government and implemented by the ANR (referenced ANR-10-LABX-54 MEMO LIFE and ANR-11-IDEX-0001-02 PSL* Research University), a start-up grant from the City of Paris and a start-up grant from the FRM. A.A.J. and A.-I.L. received fellowships from the French Ministry of Higher Education and Research.

Author Contributions A.A.J. performed and analysed experiments; A.-I.L. performed videomicroscopy experiments; N.D. performed and analysed cilia mutant experiments; M.F. contributed to *in vitro* experiments; N.S. designed the study and supervised the project; A.M. initiated the study, designed, performed and analysed experiments, and supervised the project; A.M., N.S. and A.A.J. wrote the manuscript. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.S. (nathalie.spassky@ens.fr) or A.M. (alice.meunier@ens.fr).

METHODS

Animals. All animal studies were performed in accordance with the guidelines of the European Community and French Ministry of Agriculture and were approved by the Direction départementale de la protection des populations de Paris (Approval number Ce5/2012/107). The mice used in this study have already been described and include: OF1 (Oncins France 1, Charles River Laboratories); Cen2-GFP¹⁴ (CB6-Tg(CAG-EGFP/CETN2)3-4Jgg/J, The Jackson Laboratory); *Kif3a^{fl}*, *Kif3a^{ko/+31}*; and nestin-Cre. To produce the conditional mutant *Kif3a* (*Kif3A* cKO), we crossed *Kif3a^{fl/fl}*/Cen2-GFP mice with *Kif3a^{ko/+}*/nestin-Cre mice. *Kif3a^{fl/+}*/nestin-Cre mice were used as controls and *Kif3a^{ko/fl}*/nestin-Cre mice as mutants. Videomicroscopy and 3D-SIM experiments were performed with heterozygous Cen2-GFP mice. Other experiments were performed in parallel using OF1 and heterozygous Cen2-GFP mice; no differences between these two strains were observed regarding ependymal differentiation, amplification stages, centrosome asymmetry or number of deuterosomes and centrioles.

Brain and trachea dissections. Whole mounts of developing ventricular walls were prepared from P0–adult mice as previously described³². Developing tracheae were separated surgically from the larynx (cranially) and the bronchi (caudally) of anaesthetized E17–P0 mice. For whole-mount preparations, trachea-associated muscles, vessels, nodes and glands were scraped off in cold L15 medium before cutting the trachea in half cranio-caudally with micro scissors.

Primary ependymal cell cultures and transfections. Newborn mice (P0–P2) were killed by decapitation. The brains were dissected in Hank's solution (10% HBSS, 5% HEPES, 5% sodium bicarbonate, 1% penicillin/streptomycin (P/S) in pure water) and the extracted ventricular walls were cut manually into pieces, followed by enzymatic digestion (DMEM glutamax, 33% papain (Worthington 3126), 17% DNase at 10 mg ml⁻¹, 42% cysteine at 12 mg ml⁻¹) for 45 min at 37 °C in a humidified 5% CO₂ incubator. Digestion was stopped by addition of a solution of trypsin inhibitors (Leibovitz Medium L15, 10% ovomucoid at 1 mg ml⁻¹, 2% DNase at 10 mg ml⁻¹). The cells were then washed in L15 and resuspended in DMEM glutamax supplemented with 10% fetal bovine serum (FBS) and 1% P/S in a Poly-L-lysine (PLL)-coated flask. Ependymal progenitors proliferated for 5 days until confluence followed by shaking (250 rpm) overnight. Pure confluent astroglial monolayers were replated at a density of 7×10^4 cells per cm² (corresponding to days *in vitro* (DIV) –1) in DMEM glutamax, 10% FBS, 1% P/S on PLL-coated coverslips for immunocytochemistry experiments, Lab-Tek chambered coverglasses (Thermo Fisher Scientific) for time-lapse experiments or glass-bottomed dishes with imprinted 50 µm relocation grids (Ibidi, catalogue no. 81148, Biovalley) for correlative 3D-SIM/EM and maintained overnight. The medium was then replaced by serum-free DMEM glutamax 1% P/S, to trigger ependymal differentiation gradually *in vitro* (DIV 0). All primary cultured cells derived from *Kif3a^{ko/fl}*/nestin-Cre were devoid of primary and motile cilia. Transfections were performed in suspension at DIV –1 using the Jetprime (Polyplus) system.

Plasmids. The full-length complementary DNA for Deup1 (GenBank accession number KC211186) was amplified by PCR from DIV 5 ependymal cell cultures. To express a GFP-tagged version of Deup1, the PCR fragments containing the Deup1 cDNA were cloned into pEGFP-C1. To generate a Kusabira-Orange-Deup1, the EGFP sequence of pEGFP-C1 was replaced by the Kusabira-Orange sequence using the megaprimer technique.

Immunostaining. Lateral brain ventricles and cell cultures were fixed in methanol at –20 °C for 5 min. Tracheae were permeabilized in 0.5% Triton X-100 1× PBS for 3 min before methanol fixation at –20 °C for 7 min. Tissues and cells were pre-blocked in 1× PBS with 0.2–0.5% Triton X-100 and 10% FBS before incubation with primary and secondary antibodies. Tissues and cells were counterstained with DAPI (10 µg ml⁻¹, Sigma) and mounted in Fluoromount (Southern Biotech). The following antibodies were used: rabbit anti-Ccdc78 (1:50; Sigma-Aldrich, AV53233 and HPA041186); rat anti-CD24 (1:100; BD Pharmingen); mouse IgG2b anti-Centrin 20H5 (1:2000; Millipore); rabbit anti-Cep120 (1:2000)³³; rabbit anti-Cep152 (1:2000)⁷; rabbit anti-Cep164 (1:750)³⁴; rabbit anti-CP110 (1:1000, Bethyl laboratories); rabbit anti-Cpap (1:1000)³⁵; rabbit anti-Deup1 (1:2000)⁷; mouse IgG1 anti-FoxJ1 (1:700, eBioscience); mouse IgG1 anti-GT335 (1:2000, Adipogen)³⁶; mouse IgG1 or rabbit anti-Ki67 (1:10, BD Biosciences or 1:50, Spring Bioscience); human or rabbit anti-ninein (1:100, 2g5 clone, Recombinant antibody platform, Curie Institute or 1:10000³⁷); rabbit anti-Plk4 (1:1000)³⁸; rabbit anti-Poc5 (1:500)²³; mouse IgG2b anti-Sas-6 (1:750, Santa Cruz); rabbit anti-Stil (1:1000, Bethyl Laboratories); and species-specific Alexa Fluor secondary antibodies (1:400, Invitrogen).

Microscopy. *Epifluorescence microscopy.* Fixed cells and whole-mount ventricles were examined with an upright epifluorescence microscope (Zeiss Axio Observer.Z1) equipped with an Apochromat ×63 (NA 1.4) oil-immersion objective and a Zeiss Apotome with an H/D grid. Images were acquired using Axiovision with 240-nm z-steps.

Videomicroscopy. Cells were filmed *in vitro* using an inverted spinning disk Nikon Ti PFS microscope equipped with oil-immersion ×63 (NA 1.32) and ×100 (NA 1.4)

objectives, an Evolve EMCCD Camera (Photometrics), dpss lasers (491 nm, 561 nm), appropriate filter sets for DAPI/FITC/TRITC, a motorized scanning deck and an incubation chamber (37 °C; 5% CO₂; 87% humidity). To decrease photo-toxicity, laser intensities were decreased to 3% and the image capture time was increased to 1 s per frame. Images were acquired with Metamorph Nx at different time intervals ($\Delta t = 20$ to 40 min). Image stacks were recorded with a z-distance of 0.3 µm to 0.7 µm and subjected to constrained iterative deconvolution (conservative ratio, 1–6 cycles, medium noise filtering, ImageJ software). After image acquisition, to maintain the same cell coordinates for CD24 staining, the medium was removed with a Pasteur pipette without moving the chambered coverglass, and CD24 primary and secondary antibodies were added for 15 min each, in medium supplemented with FBS (10%). The cells were then fixed for 5 min with 0.5% paraformaldehyde (PFA) before the final images were acquired (see Extended Data Fig. 2b). Four-dimensional (x, y, z, t) time-lapse images were analysed with Imaris (Bitplane Scientific Solutions) and Image J.

Transmission electron microscopy. Cultured cells or P4 ventricle slices were fixed in 2.5% glutaraldehyde and 4% PFA, treated with 1% OsO₄, washed and progressively dehydrated. The samples were then incubated in 1% uranyl acetate in 70% methanol, before final dehydration, pre-impregnation with ethanol/epon (2/1, 1/1, 1/2) and impregnation with epon resin. After mounting in epon blocks for 48 h at 60 °C to ensure polymerization, ultra-thin sections (70 nm) were cut on an ultramicrotome (Ultracut E, Leica) and analysed using a Philips Technai 12 transmission electron microscope.

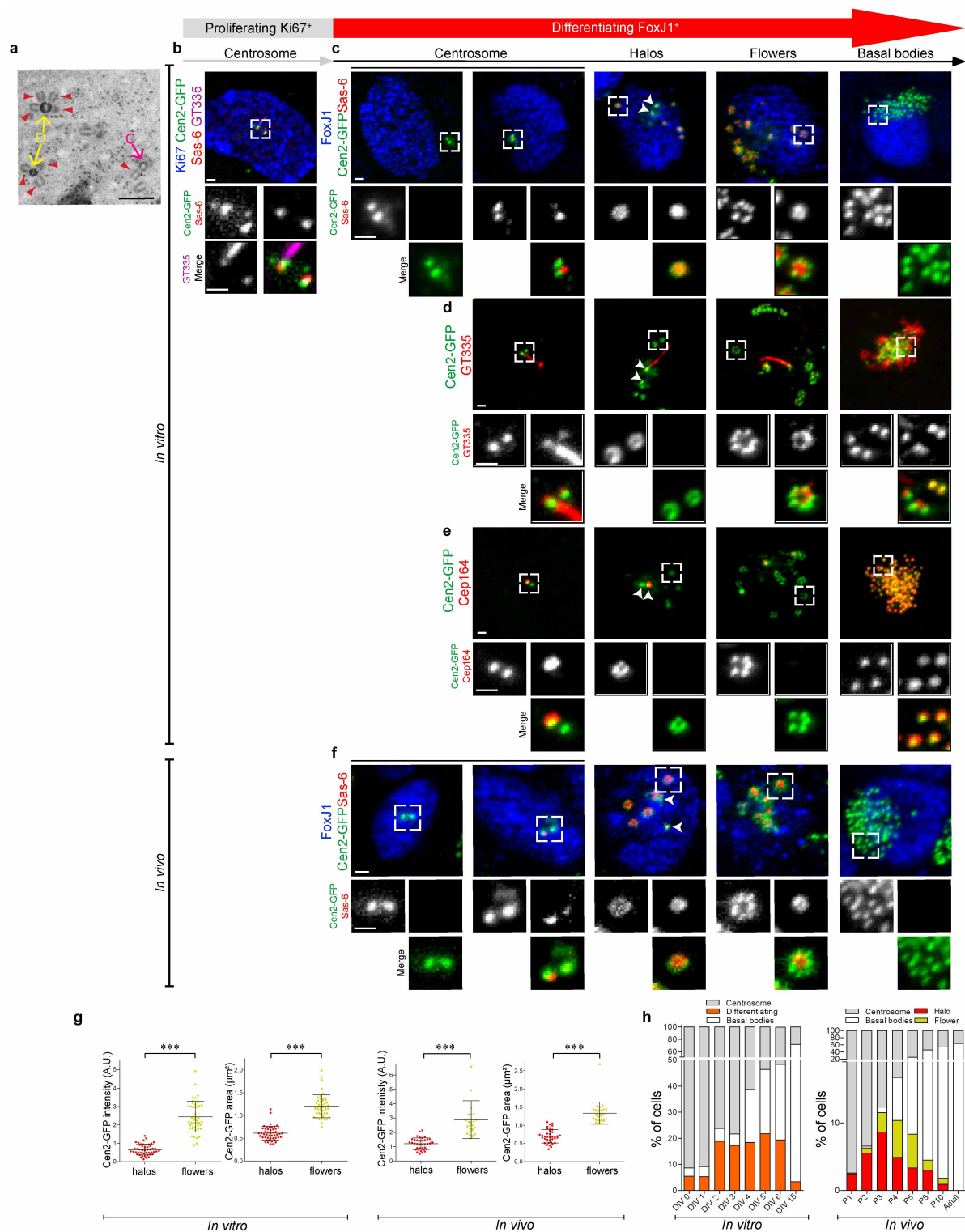
3D structured illumination microscopy. 3D-SIM images were obtained using a Nikon N-SIM Eclipse Ti Inverted Microscope equipped with a MCL Piezo stage, an Apochromat ×100 (NA 1.49) oil-immersion objective and 488/561/647 nm diode lasers. Image stacks were recorded with a z-distance of 100–120 nm, and 3D reconstruction and alignment was performed on an NIS-Elements AR microscope. Conventional images, when shown, were acquired after 3D-SIM acquisitions (see Fig. 2b). Cen2-GFP measurements were made with Image J on single 100 nm axial slices (details in Extended Data Fig. 7b).

Correlative 3D-SIM and electron microscopy. Primary Cen2-GFP ependymal progenitors were grown in 0.17-mm thick glass dishes with imprinted 50 µm relocation grids (Ibidi). At 3 days *in vitro* (DIV 3), cells were fixed with 4% PFA for 10 min and ependymal progenitors undergoing halo formation were imaged for Cen2-GFP and DAPI, in PBS, with a 3D structured illumination microscope (N-SIM - Eclipse Ti Inverted Microscope). Coordinates on the relocation grid of the cells of interest were recorded. Cells were treated for transmission electron microscopy within 2 h and embedded in epon resin. After epon polymerization, resin blocks were detached from the glass dish by several baths in liquid nitrogen. Using the grid pattern imprinted in the resin, 50 serial ultra-thin 70-nm sections of the squares of interest were cut and transferred onto formvar-coated EM grids (0.4 × 2 mm slot). The central position of the square of interest and DAPI staining are used to relocate the cell of interest (See Extended Data Fig. 3e).

Quantification and statistical analyses. Data were obtained from at least three independent experiments and the results presented as the mean ± s.d., unless otherwise stated. Non-parametric two-tailed Mann–Whitney U-tests were used to compare two distributions and two-tailed Kruskal–Wallis tests to compare groups of data, unless otherwise stated. Variances were similar between compared distributions. Distributions were tested for normality when necessary. Pearson's correlation coefficient was calculated to determine the strength of the relationship between the volume of the deuterosomes and the number of procentrioles they hold (Extended Data Fig. 6d, e). Fluorescence intensity ratios (Extended Data Fig. 1g) were obtained by dividing raw integrated densities (RID) of halos or flowers by the average RID of centrosomal centrioles (internal fluorescence control) in the corresponding cell. Daughter:mother centriole signal ratios (centrosomes in Fig. 2d, e; Extended Data Figs 2g, 3b, c and 9c, d, h, i) were calculated from centrosomes in halo-stage FoxJ1⁺ cells containing at least two cytoplasmic halos. In 15–42% of the centrosomes analysed in cells at the halo stage, Sas-6, Cep152, Plk4, Ccdc78 and Deup1 immunoreactivity was observed exclusively at the daughter centriole; these cells were excluded from quantification of the daughter:mother centriole signal ratio. Fluorescence signal intensity was quantified using Image J. All graphs and statistical analyses were obtained using GraphPad Prism software. See corresponding figure legends for P values and further details.

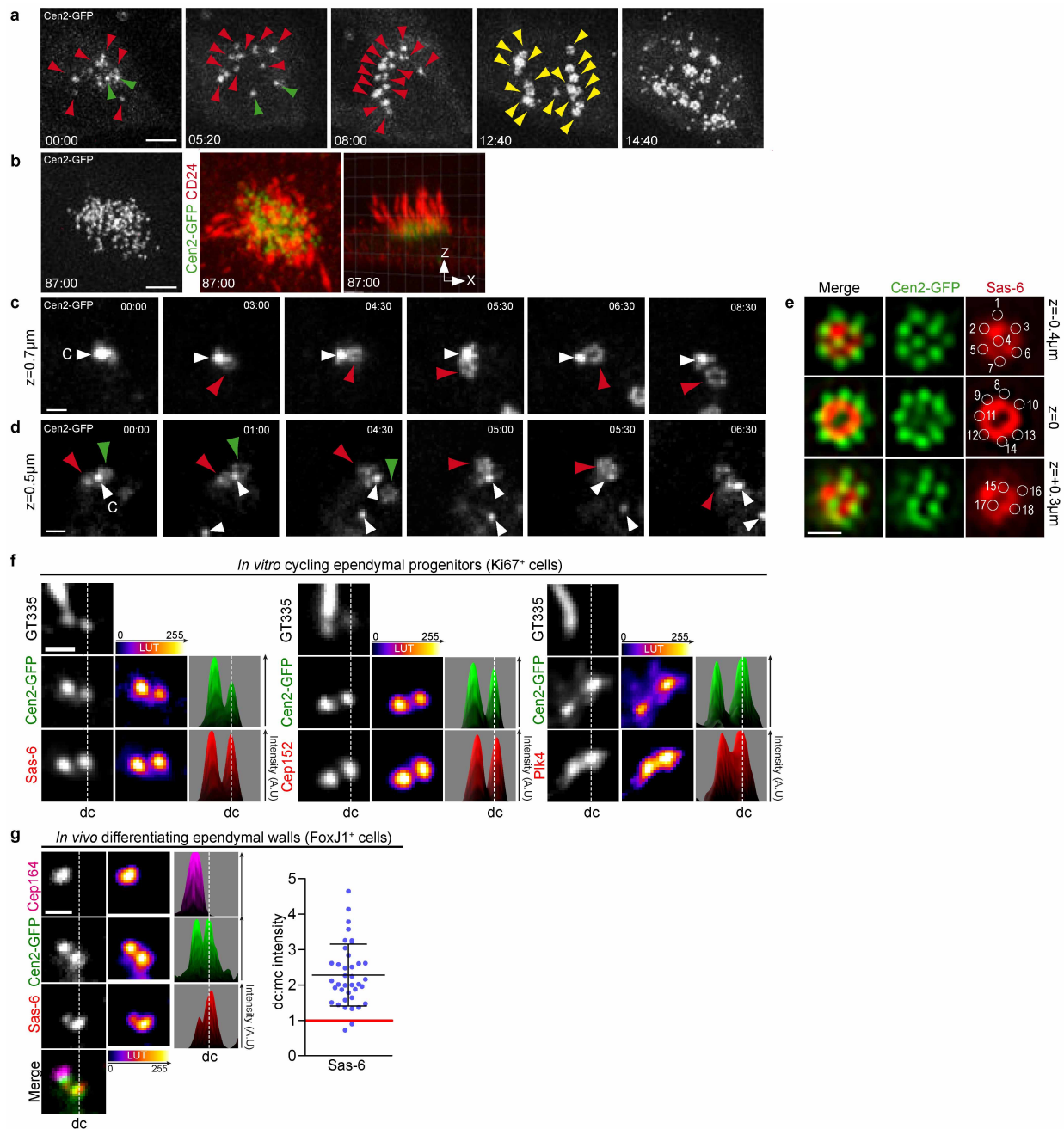
- Marszałek, J. R., Ruiz-Lozano, P., Roberts, E., Chien, K. R. & Goldstein, L. S. Situs inversus and embryonic ciliary morphogenesis defects in mouse mutants lacking the KIF3A subunit of kinesin-II. *Proc. Natl Acad. Sci. USA* **96**, 5043–5048 (1999).
- Mirzadeh, Z., Doetsch, F., Sawamoto, K., Wichterle, H. & Alvarez-Buylla, A. The subventricular zone en-face: wholemount staining and ependymal flow. *J. Vis. Exp.* **39**, 1938 (2010).
- Lin, Y. N. et al. CEP120 interacts with CPAP and positively regulates centriole elongation. *J. Cell Biol.* **202**, 211–219 (2013).
- Graser, S. et al. Cep164, a novel centriole appendage protein required for primary cilium formation. *J. Cell Biol.* **179**, 321–330 (2007).

35. Hung, L. Y., Tang, C. J. & Tang, T. K. Protein 4.1 R-135 interacts with a novel centrosomal protein (CPAP) which is associated with the γ -tubulin complex. *Mol. Cell. Biol.* **20**, 7813–7825 (2000).
36. Bosch Grau, M. *et al.* Tubulin glycyloses and glutamylases have distinct functions in stabilization and motility of ependymal cilia. *J. Cell Biol.* **202**, 441–451 (2013).
37. Delgehyr, N., Sillibourne, J. & Bornens, M. Microtubule nucleation and anchoring at the centrosome are independent processes linked by ninein function. *J. Cell Sci.* **118**, 1565–1575 (2005).
38. Kleylein-Sohn, J. *et al.* Plk4-induced centriole biogenesis in human cells. *Dev. Cell* **13**, 190–202 (2007).



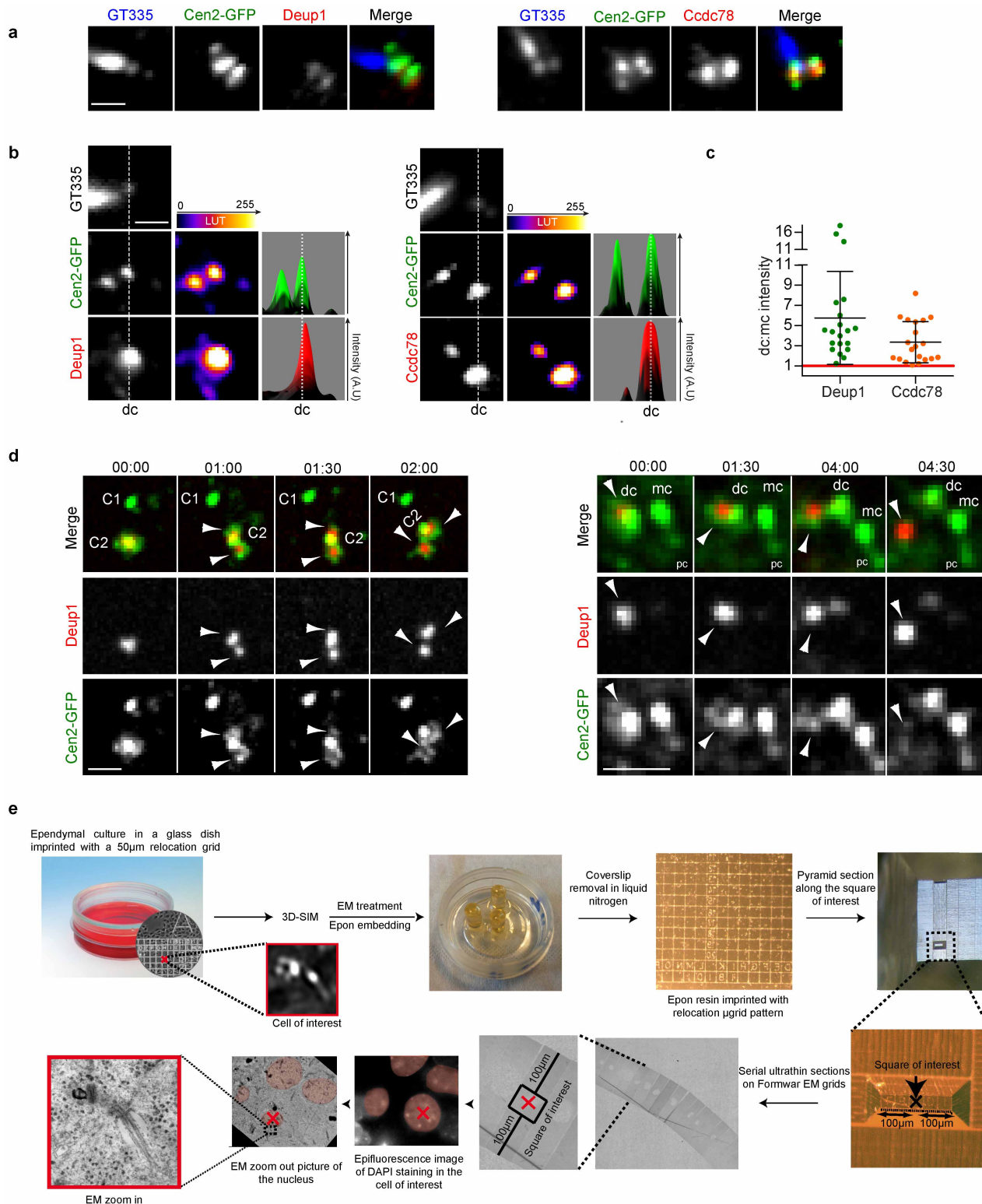
Extended Data Figure 1 | Centriole amplification in ependymal progenitor cells *in vitro* and *in vivo*. **a**, TEM showing the apical surface of a differentiating ependymal progenitor *in vitro* (DIV 3). Procentrioles (arrowheads) are seen growing from deuterosomes (D; 'deuterosome' pathway) or from a centriolar structure (C; 'centriolar' pathway). **b**, Ki67, Sas-6 and GT335 immunoreactivity during centrosome duplication in Cen2-GFP⁺ cycling ependymal progenitors *in vitro* (DIV -2). **c–e**, FoxJ1, Sas-6, GT335 and Cep164 immunoreactivity in Cen2-GFP⁺ differentiating ependymal progenitors *in vitro* (DIV 2 to DIV 5) in order of appearance (see **h**). **f**, *In vivo* immunostainings of Sas-6 and FoxJ1 in P3 to P5 Cen2-GFP⁺ ventricular walls showing that the same stages of amplification exist *in vivo* as *in vitro* (see also **h**). Boxes in **b–f** indicate

magnified structures. White arrowheads indicate the centrosomal centrioles at the halo stage. **g**, Analysis *in vitro* (DIV 3) and *in vivo* (P3–P5 lateral ventricular walls) of Cen2-GFP fluorescence showing an increase in the intensity and area of Cen2-GFP 'flower' fluorescence compared to Cen2-GFP 'halos' (*in vitro*: $n = 48$ halos; $n = 47$ flowers; *in vivo*: $n = 37$ halos; $n = 27$ flowers); error bars represent mean \pm s.d.; P values derived from two-tailed Mann-Whitney U -tests, *** $P < 0.0001$. **h**, Left: percentages of differentiating cells *in vitro* according to DIV ($n = 1,244$ cells from 3 experiments per day). Right: percentages of differentiating cells *in vivo* in the region of the lateral brain ventricle indicated by the red square in Fig. 1a according to post-natal day (P) ($n = 4,025$ cells from 2 to 3 animals per age). Scale bars: 1 μ m.



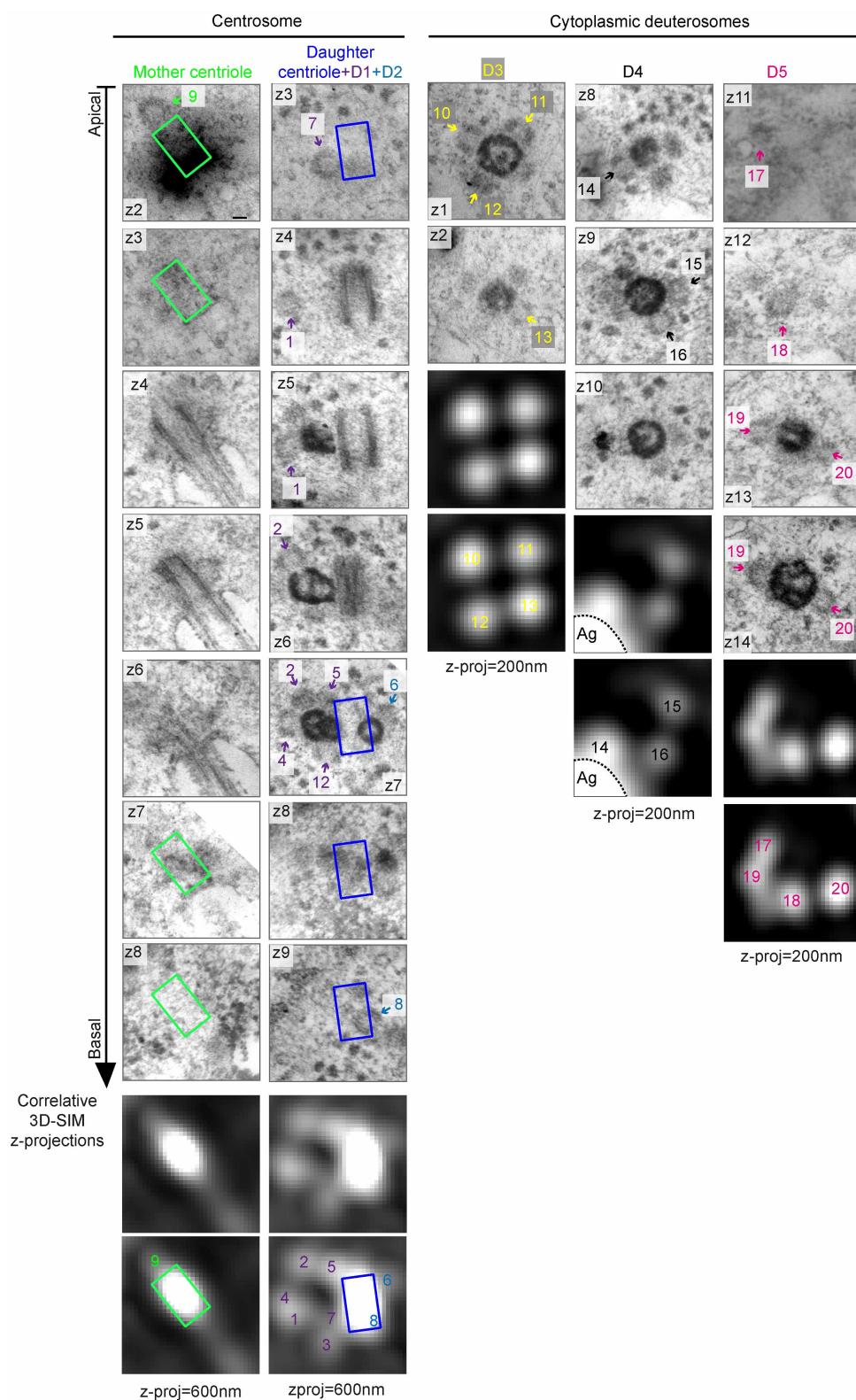
Extended Data Figure 2 | Dynamics of centriole amplification analysed live *in vitro*; centrosomes of cycling endodermal progenitors *in vitro*; centrosome Sas-6 asymmetry *in vivo*. **a**, From Supplementary Video 2. Dynamics of centriole amplification in a differentiating Cen2-GFP⁺ endodermal progenitor. Note that the video begins when some halos have already accumulated in the cytoplasm (00:00). Green arrowheads point to the centrosomal centrioles when discernible. Halos (red arrowheads) first accumulate in the cytoplasm (00:00–08:00) then transform into flowers (12:40; yellow arrowheads) before simultaneous centriole release (14:40). All the centrioles then migrate to the apical membrane where they grow cilia (87:00; **b**). **b**, Immunostaining, at the end of the time-lapse experiment shown in **a**, of the endodermal-specific membrane protein CD24 confirms that Cen2-GFP⁺ structures are docked centrioles (basal bodies) growing motile cilia. **c**, From Supplementary Video 4. Single z-plane images of a Cen2-GFP⁺ centrosomal centriole (white arrowhead) during the formation of a halo (red arrowhead). Note that the second centrosomal centriole is outside the field of observation. **d**, Single z-plane of centrosomal centrioles (white arrowheads) during the concomitant formation of two halos (red and green arrowheads) from one of the two centrosomal centrioles. The halo on the right (green arrowhead) detached at 04:30 then disappeared from the field. Note that the second centrosomal

centriole appears in the field at 01:00 (second white arrowhead). **e**, From Supplementary Video 5. Three 3D-SIM z-planes (0.1 μm) showing a cytoplasmic halo with 20 procentrioles organized in a raspberry-like structure (18 Sas-6⁺/Cen2-GFP⁺ procentrioles are numbered; circles show positions of Cen2-GFP⁺ subunits relative to Sas-6 staining). **f**, Representative images of the quantifications shown in Fig. 2e (Ki67⁺; left). Sas-6, Cep152 and Plk4 immunoreactivity on Cen2-GFP⁺ mother (GT335⁺ primary cilium) and daughter centrosomal centrioles in cycling Ki67⁺ endodermal progenitors at DIV -2. 'Fire' LUT and fluorescence intensity profiles are shown. **g**, Daughter:mother centriole Sas-6 signal ratios in FoxJ1⁺ differentiating endodermal progenitors in P4 Cen2-GFP endodermal walls. Left: Sas-6 staining on mother (Cep164⁺) and daughter (Cep164⁻) centrosomal centrioles in halo-stage cells, their corresponding 'fire' LUT's, and their fluorescence intensity profiles. Right: daughter:mother ratios of Sas-6 fluorescence intensity (error bars represent mean \pm s.d.; 2.3 ± 0.9 ; $n = 38$ cells). The distribution differs significantly from a theoretical mean ratio of 1 (one sample *t*-test, $P < 0.0001$); distributions of Sas-6 ratios in differentiating progenitors *in vitro* (Fig. 2e) and *in vivo* do not differ statistically. Time in hh:mm; scale bars, 5 μm (**a–b**); 1 μm (**c–g**).



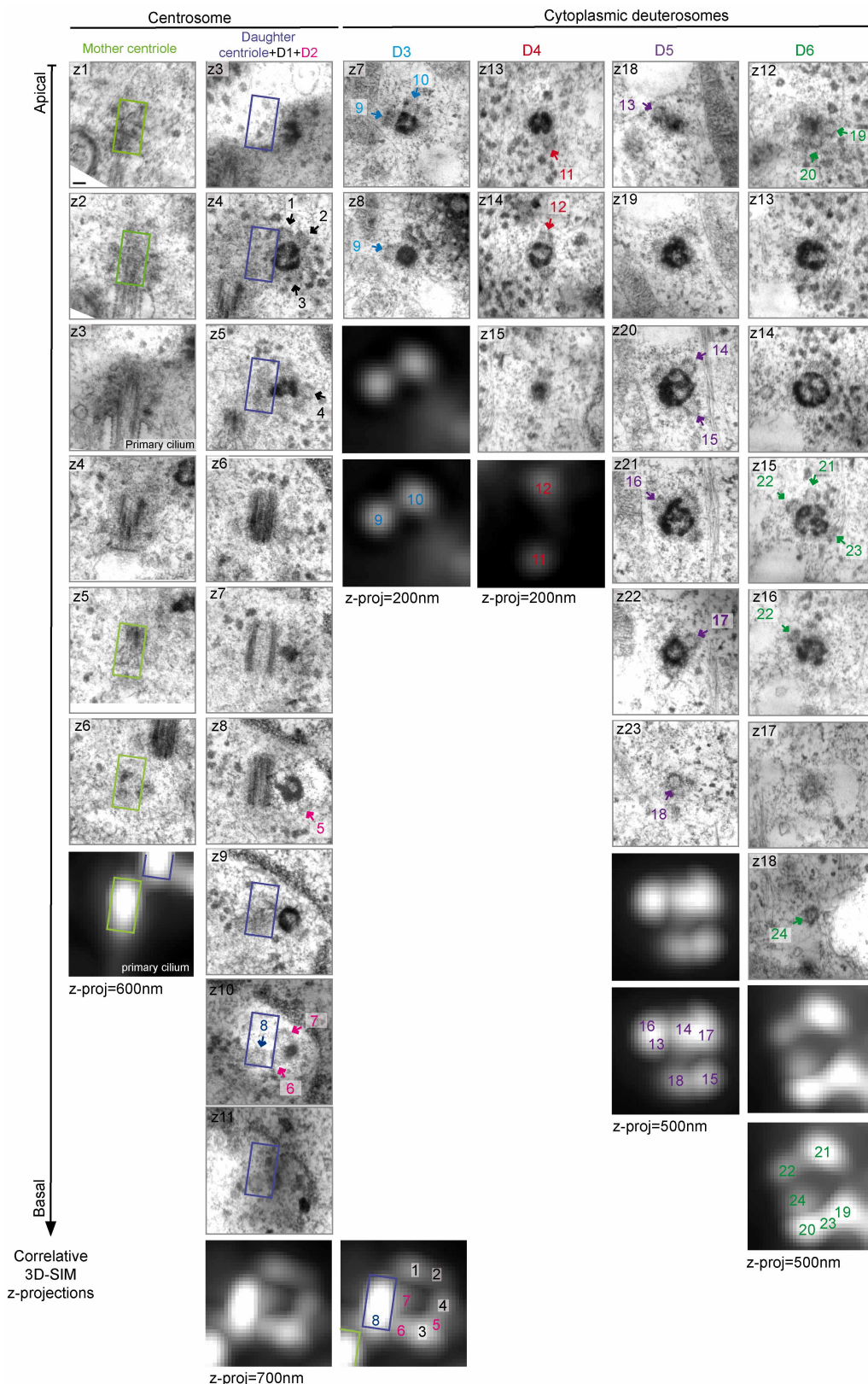
Extended Data Figure 3 | Deup1 and Ccdc78 in cycling and differentiating endymal progenitors; correlative 3D-SIM and transmission electron microscopy (3D-SIM/EM) protocol. **a**, Representative images of Deup1 and Ccdc78 immunoreactivity in cycling (Ki67⁺) Cen2-GFP⁺ endymal progenitors. **b**, Deup1 (left panel) and Ccdc78 (right panel) immunoreactivity on Cen2-GFP⁺ mother (GT335⁺ primary cilium) and daughter centrosomal centrioles in halo-stage cells. 'Fire' LUT and fluorescence intensity profiles are shown. **c**, Daughter:mother centriole Deup1 (5.7 ± 4.6 ; $n = 20$ cells) and Ccdc78 (3.4 ± 2 ; $n = 20$ cells) signal ratios in differentiating (FoxJ1⁺) endymal progenitors at the halo stage; error bars represent mean \pm s.d..

Distributions differ statistically from a theoretical mean ratio of 1 (two-tailed one sample *t*-test, $P < 0.0001$). **d**, Cen2-GFP/KusabiraOrange-Deup1 (K-O-Deup1) $\times 100$ videomicroscopy image sequences showing concentric Cen2⁺/Deup1⁺ halo formation (white arrowheads) and release from one of the two centrosomal centrioles (left panel; concomitant formation of 2 halos from C2) or from the daughter centriole (right panel; Cen2-GFP occasionally enter the primary cilium, distinguishing the mother from the daughter centriole). **e**, Overview of the correlative 3D-SIM/EM protocol (see Methods). C1 and C2: centrosomal centrioles (when mother and daughter centrioles cannot be distinguished). Time in hh:mm; scale bars, 1 µm.



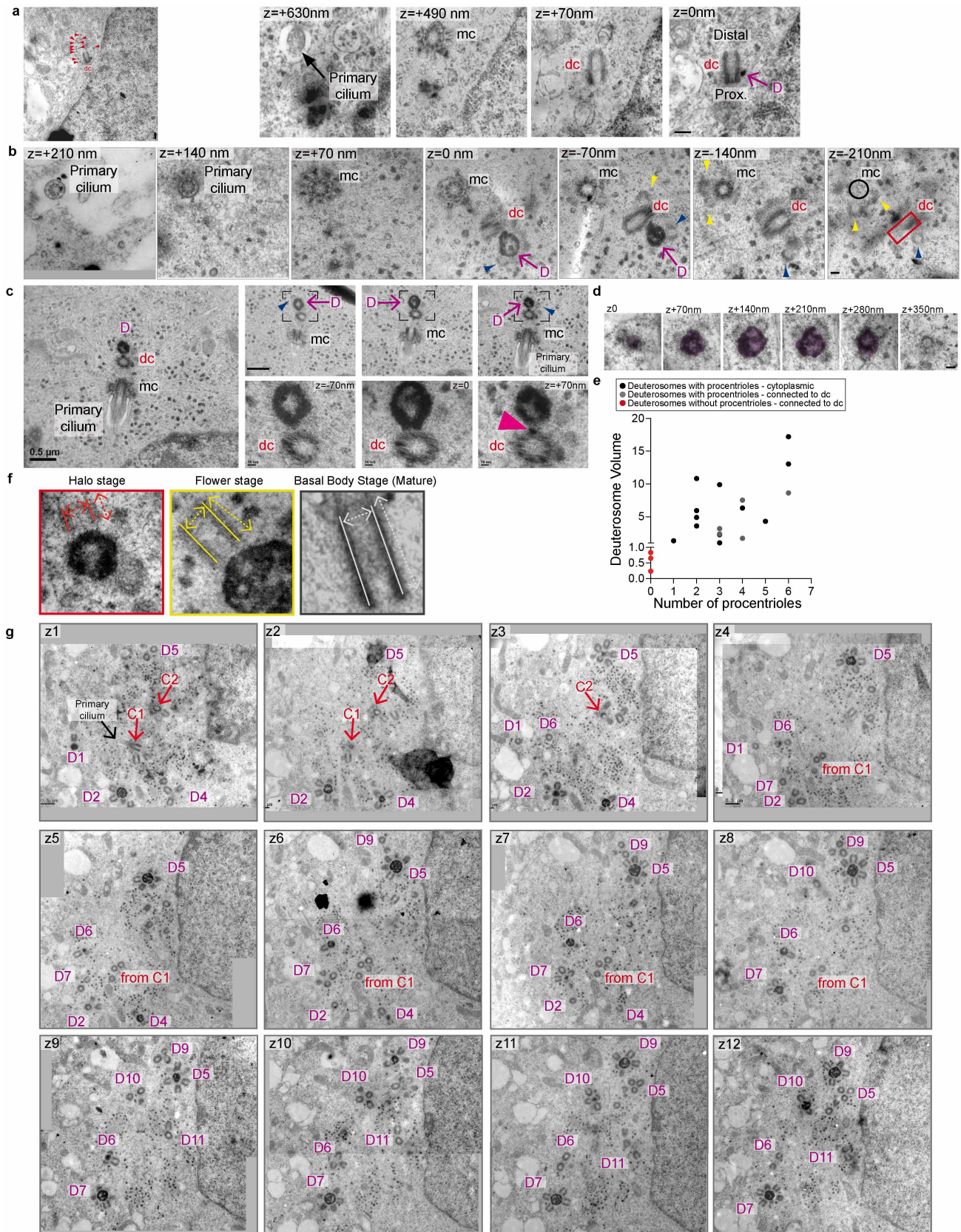
Extended Data Figure 4 | Correlative 3D-SIM and transmission electron microscopy of halo formation in a Cen2-GFP ependymal progenitor (example 1). Consecutive apical-to-basal ultra-thin EM sections (70 nm) and corresponding 3D-SIM z-projections (200 or 600 nm as indicated) are shown for both centrosomal centrioles and three cytoplasmic deuterosomes (D) in the same cell. Procentriole positions are indicated by numbers. Rectangles show the localization of the mother (green) and daughter (blue) centrioles on the

preceding or following EM section. In this cell, two deuterosomes are forming from the daughter centriole. A procentriole is growing directly from the wall of the mother centriole (9). Procentrioles cut in the sagittal plane through the centre have the same appearance (2, 4, 10, 11, 13, 14, 16). In transverse sections, the procentrioles have a tube-like structure (7, 8, 9, 18). Note that the spatial resolution of 3D-SIM is not sufficient to distinguish some of the procentrioles visible by EM (1–4, 3–7, 8). Ag, aggregate of Cen2-GFP; scale bar: 0.1 μm.



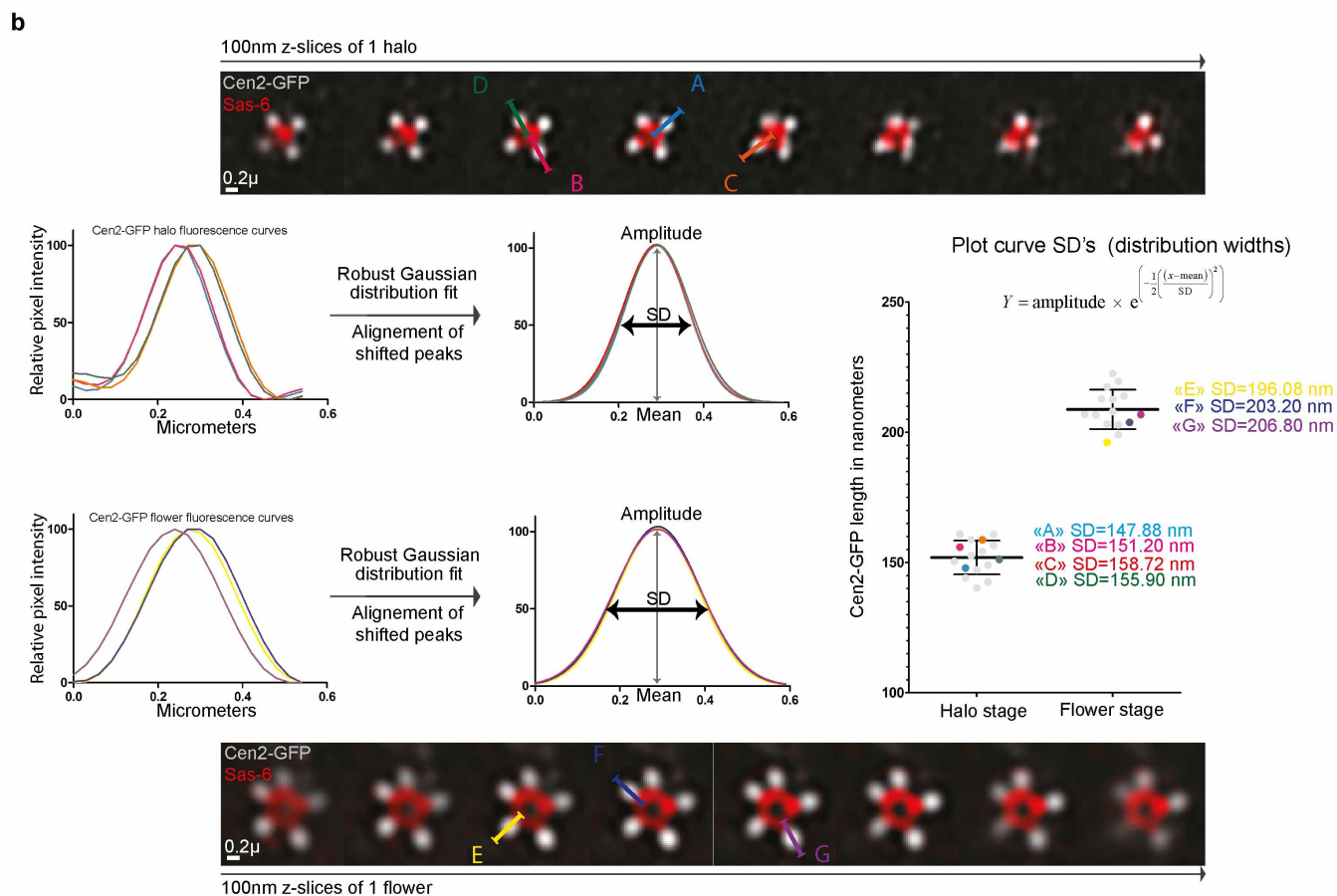
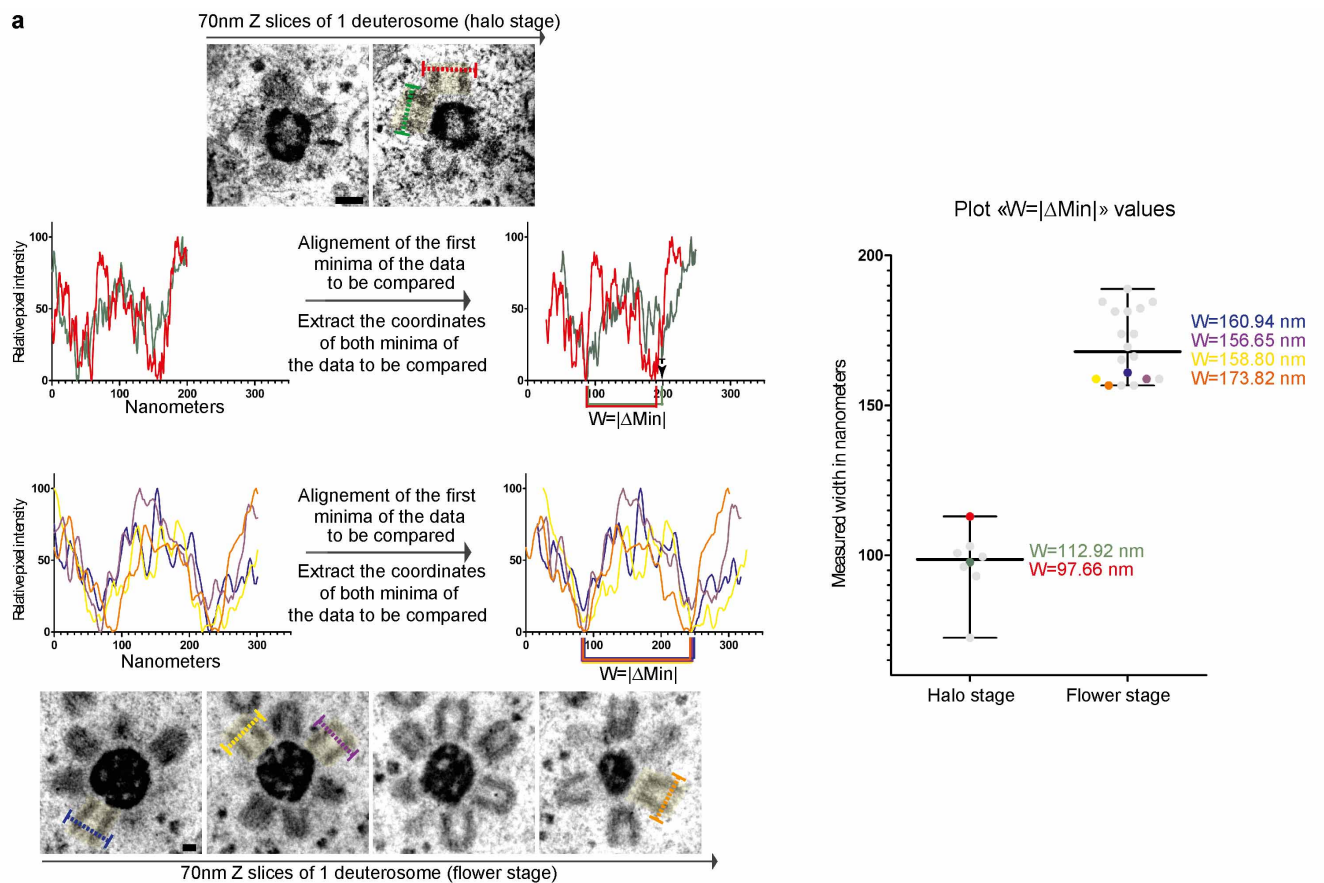
Extended Data Figure 5 | Correlative 3D-SIM and transmission electron microscopy of halo formation in a Cen2-GFP ependymal progenitor (example 2). Consecutive apical-to-basal ultra-thin EM sections (70 nm) and corresponding 3D-SIM z-projections (200–700 nm as indicated) are shown for the centrosomal centrioles and 4 cytoplasmic deuterosomes (D) in the same cell. Procentriole positions are indicated by numbers. Rectangles show the localization of the mother (green) and daughter (blue) centrioles on the

preceding or following EM section. In this cell, two deuterosomes are forming from the daughter centriole. A procentriole is growing directly from the wall of the daughter centriole (8). Procentrioles cut in the sagittal plane through the centre have the same appearance (2, 7, 15, 17, 19, 21). In transverse sections, the procentrioles have a tube-like structure (8, 18, 24). Note that the spatial resolution of 3D-SIM is not sufficient to distinguish some of the procentrioles visible by EM (3–5, 6, 8, 13–17, 19, 20–23). Scale bar, 0.1 μ m.



Extended Data Figure 6 | Formation of deuterosomes from the daughter centrosomal centriole during the halo stage and growth of procentrioles during the flower stage. **a**, Formation of the first deuterosome (D) in an early differentiating ependymal progenitor (DIV 3 cell). Left: EM image showing the daughter centrosomal centriole. In contrast to the cells presented in **b** and **c**, note the low abundance of electron-dense aggregates (red arrowheads) in the cytoplasm representative of an early step of centriologensis⁶. Right panels: serial EM sections spanning the centrosome of the same cell (note that the pictures have been rotated to re-orient the daughter centrosomal daughter centriole. Serial sections covering 2 μm of the apical portion of this cell show the absence of procentrioles and other deuterosomes. Prox., proximal. **b**, Serial EM sections spanning the centrosome of a differentiating ependymal progenitor showing a procentriole-loaded deuterosome attached to the proximal wall of the daughter centriole. Blue arrowheads point to procentrioles extending from the deuterosome. Yellow arrowheads point to procentrioles growing directly from the proximal part of the mother and the daughter centrioles. A rectangle and a circle indicate the position of the daughter and the mother centriole, respectively, in the preceding serial EM section. **c**, Left: centrosome of a cell during deuterosome formation.

Note the microtubule network converging on the mother centriole. Right panels: serial EM sections spanning the centrosome. A deuterosome is attached to the proximal part of the daughter centriole. An electron-dense patch (magenta arrowhead), previously defined as centriolar precursor material in cycling cells²⁰, can be observed on the wall of the daughter centriole adjacent to where the deuterosome contacts the centriolar wall. Blue arrowheads point to procentrioles. **d**, Deuterosome volume calculation protocol. Deuterosomes were outlined interactively (purple surfaces) on consecutive ultra-thin EM slices spanning the entire structure. Areas were summed for each deuterosome and multiplied by the thickness of the ultra-thin sections (0.07 μm). **e**, Deuterosome volume distribution relative to their localization and the number of attached procentrioles; Pearson's correlation coefficient $r = 0.67$ ($P = 0.001$) for 20 deuterosomes from 6 cells. **f**, Representative TEM images of procentrioles at the halo stage (red), the flower stage (yellow), or mature centriole docked at the plasma membrane nucleating motile cilia (grey). **g**, Serial EM sections spanning 910 nm of the apical part of a flower-stage cell (DIV 3). Procentrioles are seen growing from both centrosomal centrioles (C1 and C2) and from deuterosomes (D1 to D11). Note the raspberry-like organization of the centrioles growing from the spherical deuterosomes. Scale bars: 0.2 μm (**a**); 0.1 μm (**b**, **d**), 0.5 μm (**c**, **g**).



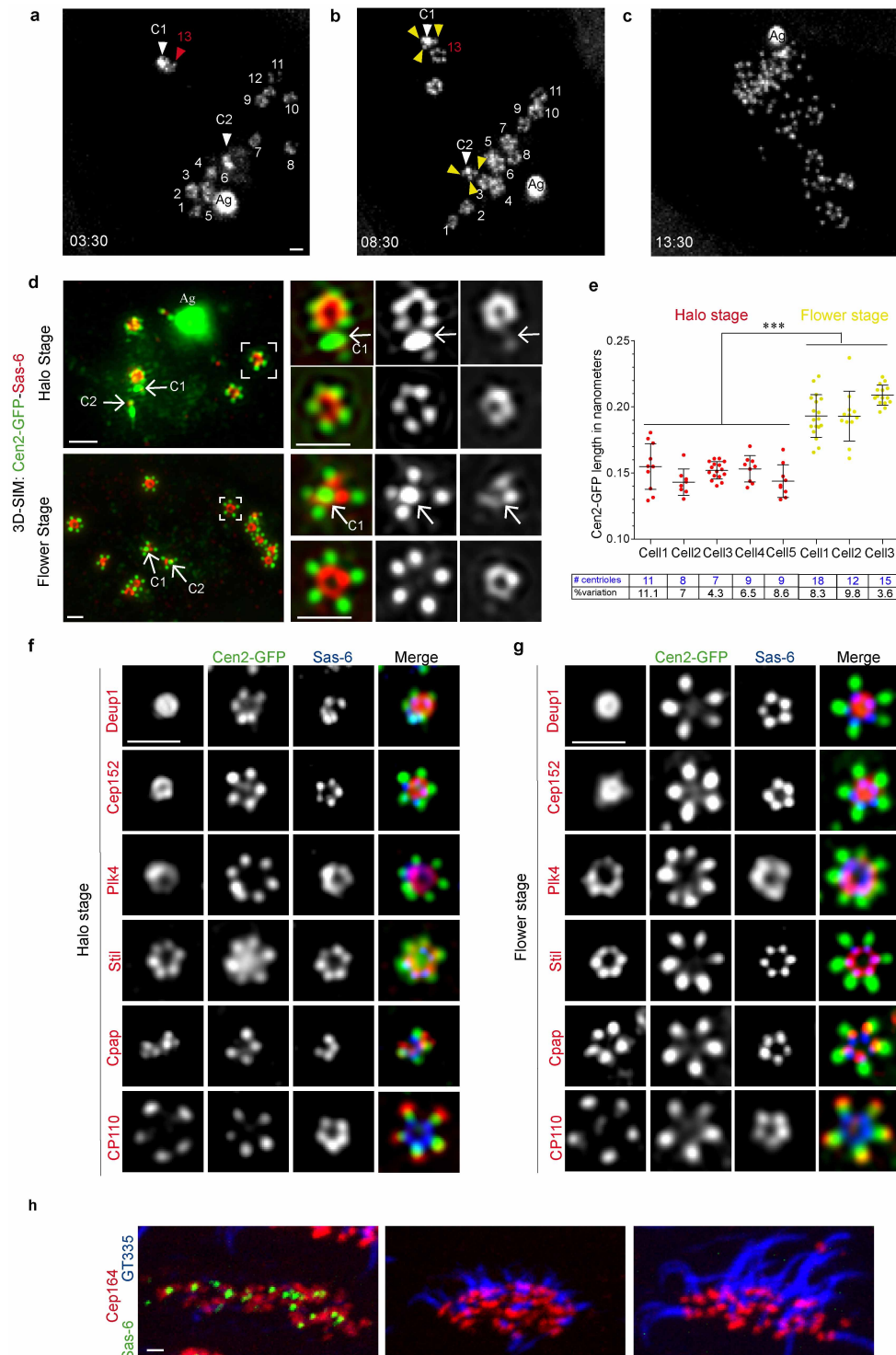
Extended Data Figure 7 | Protocols for measuring procentrioles from EM and 3D-SIM images. **a**, Protocol for measuring procentrioles from EM images.

Centrioles aligned parallel to the 70-nm z-slices were selected for quantification of EM width. Centriole width was quantified using a pixel intensity measurement tool placed at the distal extremities of the centrioles perpendicular to the centriolar walls. The resulting distributions of normalized pixel intensity were characterized by two minima on the y-coordinates corresponding to the centre of each of the two centriolar walls. Final plotted widths were obtained using the equation $W = |\Delta\text{Min}|$; where W is the absolute difference in nm between the x-coordinates of both y-axis minima.

b, Protocol for measuring procentriole lengths from 3D-SIM images. Measurements were made of Cen2-GFP procentrioles, stained with Sas-6, that

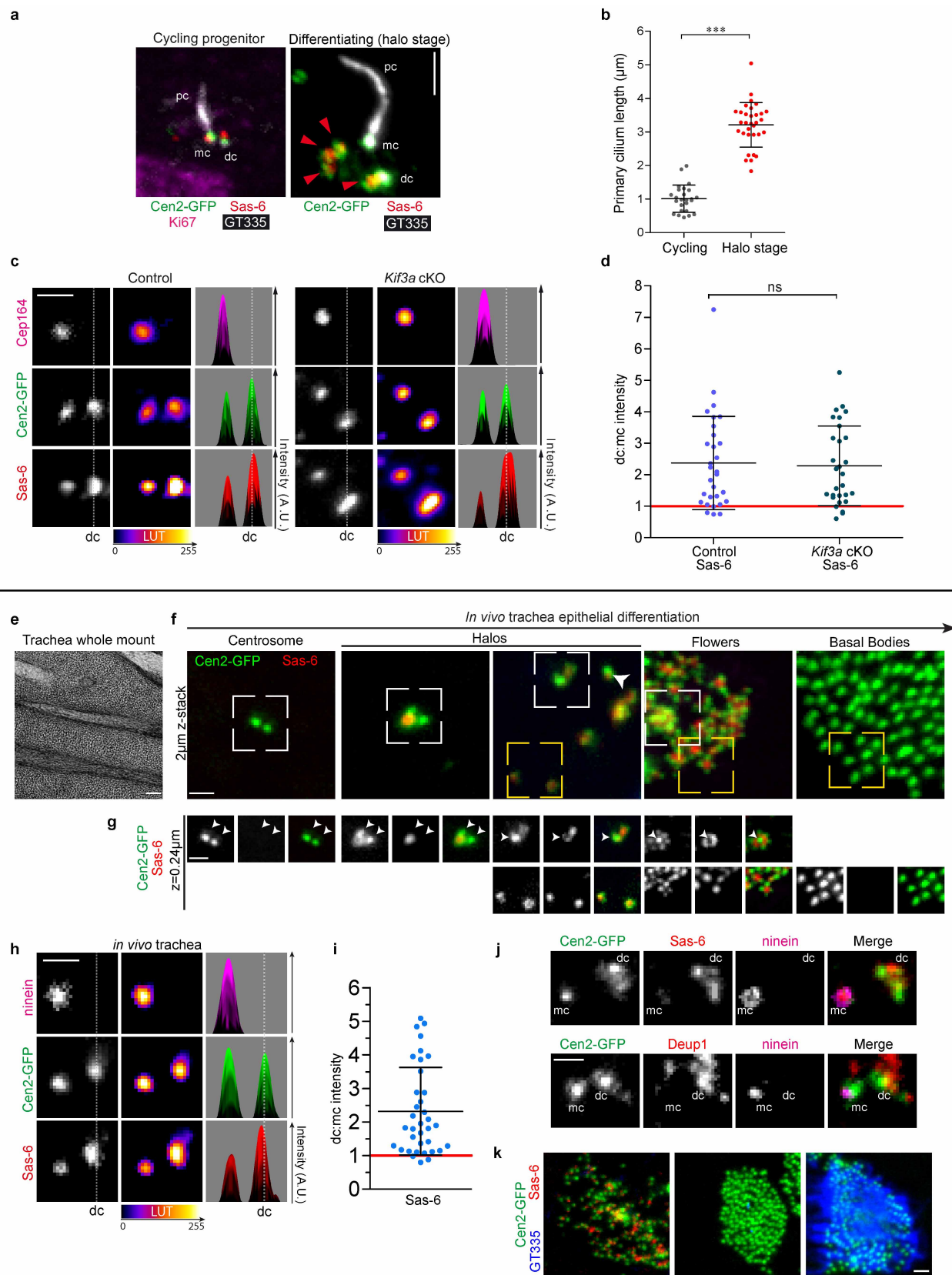
were parallel to the x,y plane on single 100-nm z-slices. Sas-6 staining was used to determine the longitudinal axis of the procentriole and was excluded from measurements of fluorescence intensity. Distributions of normalized Cen2-GFP fluorescence intensities were fitted using the robust nonlinear regression method, aligned, and their widths were measured to obtain a plot of procentriole lengths in nanometers. Each plotted dot corresponds to the width of the distribution curve obtained using the equation

$Y = \text{amplitude} \times e^{\left(-\frac{1}{2} \left(\frac{(x - \text{mean})}{\text{SD}}\right)^2\right)}$, where SD is the width of the distribution, amplitude is the height of the centre of the distribution in Y units, and mean is the x value at the centre of the distribution.



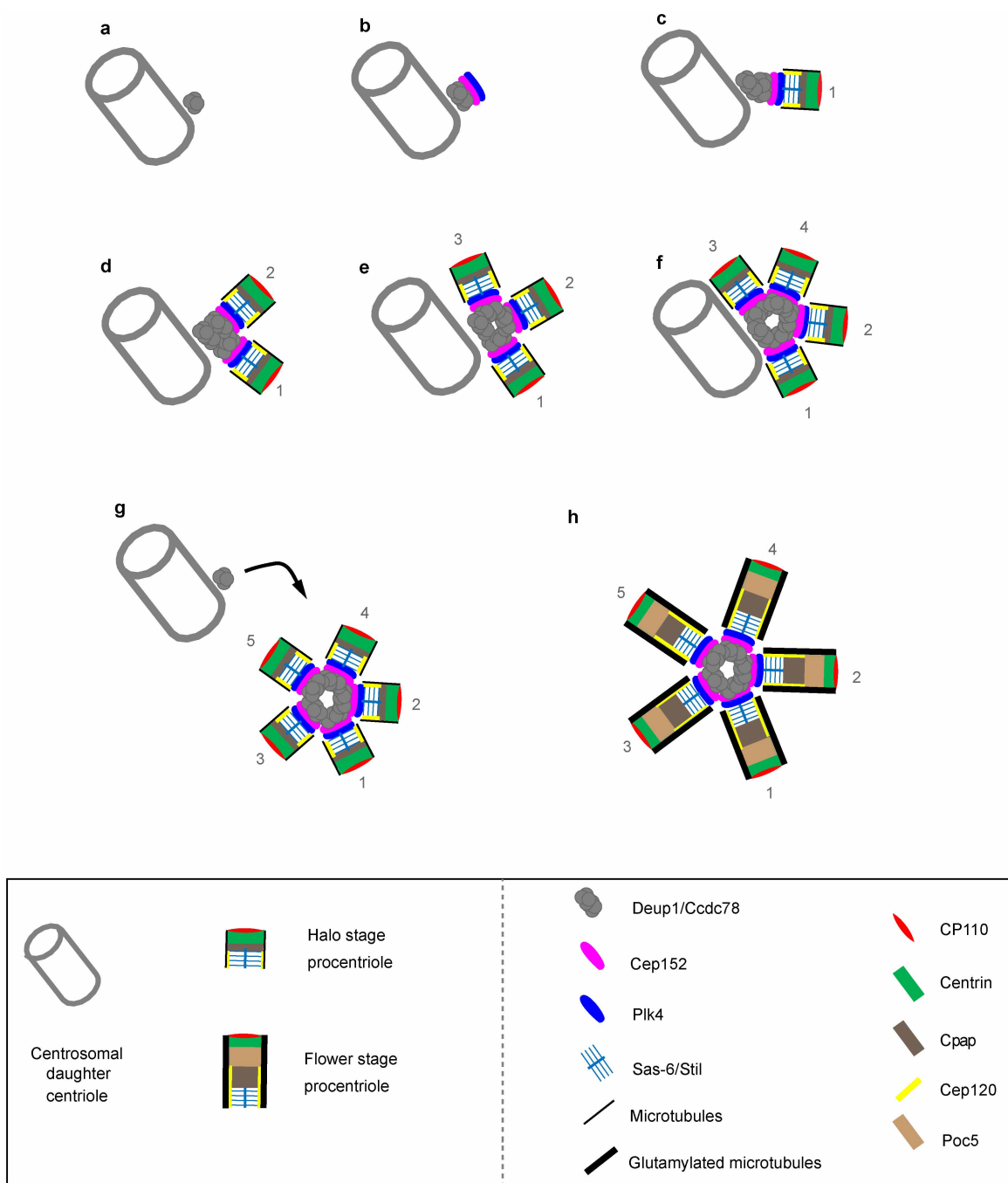
Extended Data Figure 8 | Halo to flower stage transition in differentiating ependymal progenitors. **a–c**, From Supplementary Video 4. Dynamics of the transformation of Cen2–GFP halos (numbered at 03:30) into flower-like structures (corresponding numbers at 08:30) after the formation of the thirteenth and last halo (red arrowhead, same cell as in Extended Data Fig. 2c). Note that procentrioles become visible on the walls of both centrosomal centrioles at the flower stage (yellow arrowheads). Centrioles detach simultaneously from both centrosome and deuterosome platforms at 10:30. C1 and C2 are centrosomal centrioles. Ag, aggregate of Cen2–GFP. **d**, Sas-6 immunoreactivity on 3D-SIM z-projections (left panel) or single z-slices (120 nm, right panels) of Cen2–GFP⁺ (green) cells at the halo and flower stages. At the halo stage, halos have accumulated in the cytoplasm and a halo is seen forming from the C1 centriole (same cell as in Fig. 2c). At the flower stage,

eleven flowers are visible in the cytoplasm and procentrioles are elongating from C1 and C2 centrosomal centrioles and deuterosomes. **e**, Quantification of the length of Cen2–GFP⁺ procentrioles on 3D-SIM images of cells at the halo (red) and flower (yellow) stage. There are no significant differences among cells. The lengths of the procentrioles in flower stage-cells are statistically different from the lengths of the procentrioles in halo-stage cells (error bars represent mean \pm s.d.; two-tailed unpaired *t*-test, ****P* < 0.0001). See Extended Data Fig. 7b for quantification protocol. **f, g**, 3D-SIM z-slices (120 nm) showing the immunolocalization of Sas-6, Deup1, Cep152, Plk4, Stil, Cpap and CP110 in Cen2GFP⁺ halo (**f**) and flower (**g**) procentrioles. **h**, Immunostaining of P4 ependymal walls showing that Sas-6 immunoreactivity disappears from Cep-164⁺ centrioles when they begin to grow motile cilia (GT335⁺). Time in hh:mm; scale bars, 1 μ m.



Extended Data Figure 9 | Ciliary mutant analyses *in vitro* and centriole amplification in mouse trachea *in vivo*. **a–d**, Ciliary mutant analyses *in vitro*. **a**, GT335 and Sas-6 immunoreactivity in a Cen2-GFP⁺ cycling ependymal progenitor (DIV -2; Ki67⁺) and a differentiating progenitor at the halo stage (DIV3). Arrowheads point to halos. **b**, Quantification of primary cilium length in cycling ($1 \pm 0.4 \mu\text{m}$; $n = 25$ cells) and differentiating progenitors ($3.2 \pm 0.6 \mu\text{m}$; $n = 32$ cells) showing a significant elongation at the halo stage (error bars represent mean \pm s.d.; two-tailed Mann–Whitney *U*-test, *** $P < 0.0001$). **c, d**, The absence of a primary cilium in differentiating ependymal progenitor cells from *Kif3a* cKO mutants does not affect Sas-6 asymmetry. **c**, Representative Sas-6 staining on Cen2-GFP⁺ mother (Cep164⁺) and daughter centrosomal centrioles in control (left panel) and *Kif3a* cKO (right panel) halo-stage cells. Second and third columns show, respectively, the corresponding ‘fire’ LUT signal intensities and fluorescence intensity profiles. **d**, Daughter:mother centriole Sas-6 signal ratios in differentiating (FoxJ1⁺) control (2.4 ± 1.5 ; $n = 30$ cells) and *Kif3a* cKO (2.3 ± 1.3 ; $n = 29$ cells) ependymal progenitor cells at the halo stage. The distributions are not statistically different, but both differ significantly from a theoretical mean ratio of 1 (error bars represent mean \pm s.d.; two-tailed one sample *t*-test, $P < 0.0001$; ns, not significant). **e–k**, Centriole amplification in mouse trachea *in vivo*. **e**, Bright-field image showing a typical cranio-dorsal region of the E17–P0 trachea whole mounts analysed in (f–k). **f**, Merged images of Sas-6-immunostained Cen2-GFP⁺ tracheal multiciliated progenitors showing that the steps of centriole amplification are the same as in brain

respectively in the upper and lower panels of **g**. **g**, Single $0.24 \mu\text{m}$ z-slices of the boxed regions in **f**: the centrosome images show Sas-6 negative centrosomal centrioles; the halo images show that halos arise from the wall of a centrosomal centriole (upper panels) and accumulate in the cytoplasm (lower panels); the flower images show procentrioles on a centrosomal centriole (centriolar pathway; upper panels) and hollow Sas-6⁺ flowers (deuterosome pathway; lower panels); the basal body images show mature docked Sas-6-negative basal bodies. White arrowheads point to centrosomal centrioles in **f** and **g**. **h**, Sas-6 staining on Cen2-GFP⁺ mother (ninein⁺) and daughter centrosomal centrioles in halo-stage cells in trachea epithelium. Second and third columns show, respectively, the corresponding ‘fire’ LUT signal intensities and fluorescence intensity profiles of the same markers. **i**, Daughter:mother centriole Sas-6 signal ratios (2.3 ± 1.3 ; $n = 36$ cells) in halo-stage cells in E17–P0 trachea. The distribution differs statistically from a theoretical mean ratio of 1 (error bars represent mean \pm s.d.; two-tailed one sample *t*-test, $P < 0.0001$); Sas-6 halo-stage daughter/mother centriole ratios are similar in ependymal walls (Extended Data Fig. 2g) and trachea epithelium. **j**, Upper panels: $0.96 \mu\text{m}$ z-stacked images of a typical Sas-6/ninein immunoreactivity on E17–P0 Cen2-GFP trachea showing that halo structures form exclusively on the daughter centriole ($n = 12$ cells). Lower panels: $0.96 \mu\text{m}$ z-stacked images of a typical Deup1/ninein immunoreactivity on E17–P0 Cen2-GFP trachea showing accumulation of Deup1 on the daughter centriole. **k**, Immunostaining of P0 trachea showing that Sas-6 immunoreactivity disappears from basal bodies when they begin to grow GT335⁺ motile cilia. Scale bars, $1 \mu\text{m}$.



Transcriptional regulation of autophagy by an FXR–CREB axis

Sunmi Seok^{1*}, Ting Fu^{1*}, Sung-E Choi^{1,2}, Yang Li³, Rong Zhu⁴, Subodh Kumar¹, Xiaoxiao Sun⁴, Gyesoon Yoon², Yup Kang², Wenxuan Zhong⁴, Jian Ma³, Byron Kemper¹ & Jongsook Kim Kemper¹

Lysosomal degradation of cytoplasmic components by autophagy is essential for cellular survival and homeostasis under nutrient-deprived conditions^{1–4}. Acute regulation of autophagy by nutrient-sensing kinases is well defined^{3,5–7}, but longer-term transcriptional regulation is relatively unknown. Here we show that the fed-state sensing nuclear receptor farnesoid X receptor (FXR)^{8,9} and the fasting transcriptional activator cAMP response element-binding protein (CREB)^{10,11} coordinately regulate the hepatic autophagy gene network. Pharmacological activation of FXR repressed many autophagy genes and inhibited autophagy even in fasted mice, and feeding-mediated inhibition of macroautophagy was attenuated in FXR-knockout mice. From mouse liver chromatin immunoprecipitation and high-throughput sequencing data^{12–15}, FXR and CREB binding peaks were detected at 178 and 112 genes, respectively, out of 230 autophagy-related genes, and 78 genes showed shared binding, mostly in their promoter regions. CREB promoted autophagic degradation of lipids, or lipophagy¹⁶, under nutrient-deprived conditions, and FXR inhibited this response. Mechanistically, CREB upregulated autophagy genes, including *Atg7*, *Ulk1* and *Tfeb*, by recruiting the coactivator CRTC2. After feeding or pharmacological activation, FXR *trans*-repressed these genes by disrupting the functional CREB–CRTC2 complex. This study identifies the new FXR–CREB axis as a key physiological switch regulating autophagy, resulting in sustained nutrient regulation of autophagy during feeding/fasting cycles.

Metabolic nuclear receptors function as intracellular biosensors for lipid metabolites and nutrients, including bile acids and fatty acids, and transduce the nutrient signals into transcriptional programming to maintain homeostasis^{8,9}. The nuclear receptor FXR is activated by increased bile acid levels after feeding, and controls postprandial metabolic responses^{8,9}. Because FXR is a fed-state sensing gene regulator, we postulated that FXR transcriptionally represses hepatic autophagy in the postprandial period.

From mouse liver chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) data^{12,13}, FXR binding peaks were detected near the transcription start sites (TSSs) of 178 out of 230 autophagy-related genes involved in key autophagic processes (Supplementary Table 1, selected genes displayed in Extended Data Fig. 1a). FXR, therefore, may directly repress hepatic autophagy. To confirm the ChIP-seq results, liver ChIP analysis was performed in mice treated with GW4064, a specific agonist of FXR, rather than bile acids, which activate several signalling pathways⁹. Treatment with GW4064 in fasted mice increased FXR occupancy at many autophagy genes, and decreased messenger RNA levels of most of these genes in wild-type mice, but not in FXR-knockout (*Fxr*^{−/−}) mice, and also decreased their pre-mRNA levels (Fig. 1a, b and Extended Data Fig. 1b–d). Activation of FXR, thus, increased its occupancy at many autophagy genes and repressed their expression.

We next tested whether activation of FXR suppresses macroautophagy. Fasting increased the number of autophagic vesicles in liver, but the

increase was significantly blunted by GW4064 treatment (Fig. 1c). Mitochondria in double-membrane autophagosomes or autolysosomes, known as mitophagy, were detected in fasted wild-type mice but not after GW4064 treatment. Increased abundance of LC3 conjugated to phosphatidylethanolamine (LC3-II) is an indicator of autophagic flux¹⁷, whereas the autophagosome adaptor p62 (also known as SQSTM1) accumulates when autophagy is inhibited¹⁷. Treatment with GW4064 decreased LC3-II and increased p62 levels in mouse liver extracts (Fig. 1d), and decreased LC3-II levels were observed in hepatocytes after lysosomal inhibition with the autophagy inhibitors bafilomycin A₁ (Fig. 1e) or chloroquine (Extended Data Fig. 2a). These results indicate that FXR inhibits autophagy by transcriptionally repressing autophagy components, probably independently of lysosomal function.

To avoid confounding *in vivo* effects, such as induction of intestinal FGF19 by FXR¹⁸, direct FXR effects on autophagy were examined in mouse hepatoma Hepa1c1c7 cells. The increase in green fluorescent protein (GFP)–LC3 puncta in cells incubated in HBSS starvation media was attenuated by GW4064 treatment, but not if FXR was also downregulated (Fig. 1f). These findings indicate that FXR directly inhibits autophagy.

Because FXR is activated after feeding by bile acids, we tested whether FXR regulates autophagy under physiological conditions. The decrease in GFP–LC3 puncta after feeding in wild-type mice was markedly attenuated in *Fxr*^{−/−} mice (Fig. 2a and Extended Data Fig. 2b), and feeding decreased LC3-II levels and increased p62 levels in wild-type, but not in *Fxr*^{−/−} mice (Fig. 2b). These findings indicate that FXR acts as a key physiological repressor of autophagy in the fed state.

After feeding, autophagy components are suppressed by the nutrient-sensing mammalian target of rapamycin (mTOR)–S6 kinase pathway^{3,5–7}, which could potentially mediate FXR inhibition of autophagy. Activation or downregulation of FXR resulted in expected changes in LC3-II levels, but no changes in phosphorylated S6 (p-S6) levels (Fig. 2c). Similar effects were observed after GW4064 treatment in hepatocytes (Fig. 2d) and liver extracts (Fig. 2e). Furthermore, treatment with the mTOR inhibitor rapamycin increased LC3-II levels as expected, and co-treatment with GW4064 prevented the increase without changing the p-S6 levels (Fig. 2f). These results indicate that FXR-mediated inhibition of autophagy seems to be independent of mTOR activation.

To examine how FXR inhibits autophagy genes further, we analysed DNA binding motifs in FXR binding peak regions detected by ChIP-seq¹². Inverted repeat 1 (IR1) DNA binding motifs for FXR were not detected in sequences common to the FXR peak regions, but binding sites for several transcription factors (Supplementary Table 2) were detected, including the top scoring motifs, TGACGT(C/T)(A/T), for CREB (Fig. 3a), a key transcriptional activator that drives fasting responses^{10,11}. Remarkably, in published liver ChIP-seq data for CREB^{15,19} and FXR^{12,13}, CREB peaks were detected at 112 out of 230 autophagy genes (Supplementary Table 3), and overlapped with FXR peaks in 78 genes (Fig. 3b and Supplementary Table 4), mostly within 1 kilobase (kb) of the TSS (Supplementary

¹Department of Molecular and Integrative Physiology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. ²Institute for Medical Science, Ajou University School of Medicine, Suwon 442-749, Korea. ³Department of Bioengineering and the Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. ⁴Department of Statistics, University of Georgia, Athens, Georgia 30602, USA.

*These authors contributed equally to this work.

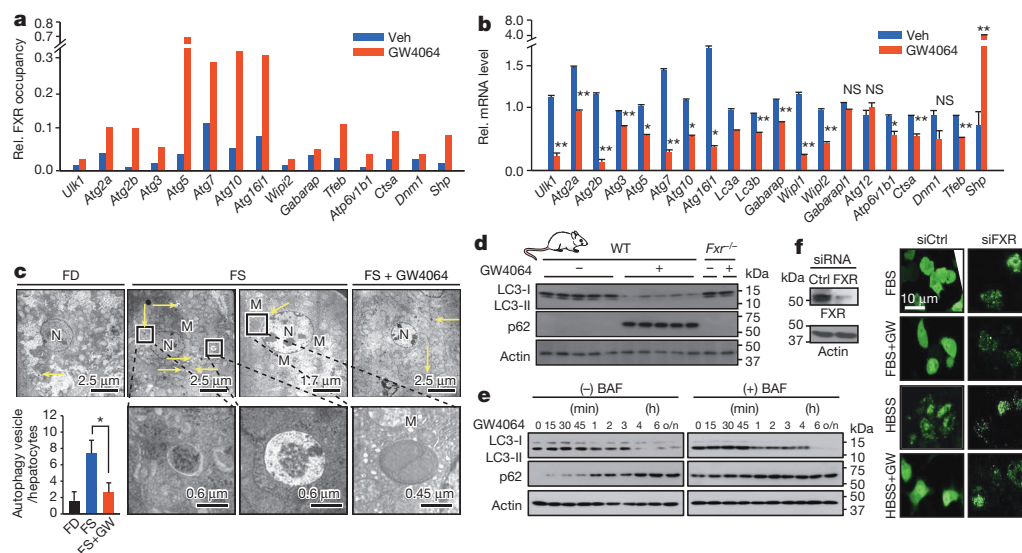


Figure 1 | Pharmacological activation of FXR transcriptionally inhibits hepatic autophagy. **a**, **b**, FXR occupancy by ChIP (**a**) and mRNA levels by qRT-PCR (**b**) in mice treated with GW4064 for 6 h after 10 h fasting ($n = 3$). *Lc3b* is also known as *Map1lc3b*; *Shp* is also known as *Nr0b2*. **c**, Electron microscopy of liver sections. Arrows indicate autophagosomes and mitochondria ($n = 10$ hepatocytes). FD, feeding; FS, fasting; M, mitochondria; N, nucleus.

d, Hepatic LC3 and p62 levels in wild-type (WT) or *Fxr*^{-/-} mice. kDa, kilodaltons. **e**, Effects of bafilomycin A₁ (BAF) and/or GW4064 on LC3 and p62 levels in hepatocytes. o/n, overnight. **f**, GFP-LC3 fluorescence in Hepa1c1c7 cells, HBSS (starvation media) and GW (GW4064). siCtrl, control short interfering RNA (siRNA); siFxr, siRNA targeting *Fxr*. Data are mean \pm s.e.m. NS, not significant. * $P < 0.05$, ** $P < 0.01$ (Student's *t*-test).

Table 3). Selected autophagy-related genes with overlapping FXR- and CREB-binding sites are listed in Fig. 3c and peaks are displayed in Extended Data Fig. 3a. Co-occupancy of FXR and CREB at these genes was confirmed in re-ChIP assays (Extended Data Fig. 3b). Downregulation of CREB decreased expression of autophagy genes (Fig. 3d) and reduced the number of GFP-LC3 puncta (Fig. 3e) in hepatic cells. These results indicate that CREB is a novel transcriptional activator of autophagy and that FXR may antagonize the CREB activity.

We next examined autophagic degradation of lipids, known as lipophagy¹⁶. Lipophagy, detected by co-staining of LC3 puncta with the fluorophore dipyrromethene boron difluoride (BODIPY 581/591), was reduced in fasted cells by downregulation of CREB (Fig. 3f). Furthermore, overexpression of CREB reversed the inhibition of autophagy mediated by downregulation of ATG7 (Extended Data Fig. 4a, b). By contrast, downregulation of FXR increased lipophagy in non-fasted cells (Fig. 3g), and FXR activation decreased lipophagy in fasted cells (Extended Data Fig. 4c). These results indicate that CREB and FXR oppositely regulate lipophagy in response to nutritional conditions.

To examine regulation of autophagy genes by FXR and CREB, an *Atg7* or *Tfeb* DNA fragment that contained FXR and CREB binding peaks was inserted into a luciferase vector (*Atg7*-luc and *Tfeb*-luc, respectively).

Overexpression of CREB and its coactivator CRTC2 (ref. 10) increased, whereas overexpression of FXR inhibited luciferase activity (Fig. 4a). Mutation of the CREB site or downregulation of CREB blocked the FXR inhibition (Extended Data Fig. 5a, b). *Tfeb*-luc activity that was increased by CREB-CRTC2 was attenuated by FXR, but not by other nuclear receptors (Extended Data Fig. 5c). In co-immunoprecipitation studies, the interaction of CREB with CRTC2 was blocked, whereas interaction of CREB with FXR was increased by GW4064 treatment (Extended Data Fig. 5d). In glutathione S-transferase (GST) pull-down studies, CREB interacted with the carboxy-terminal domain of FXR (Extended Data Fig. 5e). These results indicate that FXR may *trans*-repress autophagy genes by directly interacting with CREB and suppressing its activity. Supporting this conclusion, a DNA-binding-deficient FXR mutant did not increase FXRE-luc activity but repressed *Tfeb*-luc activity (Extended Data Fig. 5f, g).

The functional interaction of FXR with the CREB-CRTC2 complex was next examined by ChIP assay. Occupancy of CREB at autophagy genes was not changed, whereas that of FXR and CRTC2 was increased and decreased, respectively, and RXR- α occupancy was barely detectable after GW4064 treatment (Extended Data Fig. 6a, b). In re-ChIP assays, GW4064 treatment decreased CRTC2 and increased FXR occupancies

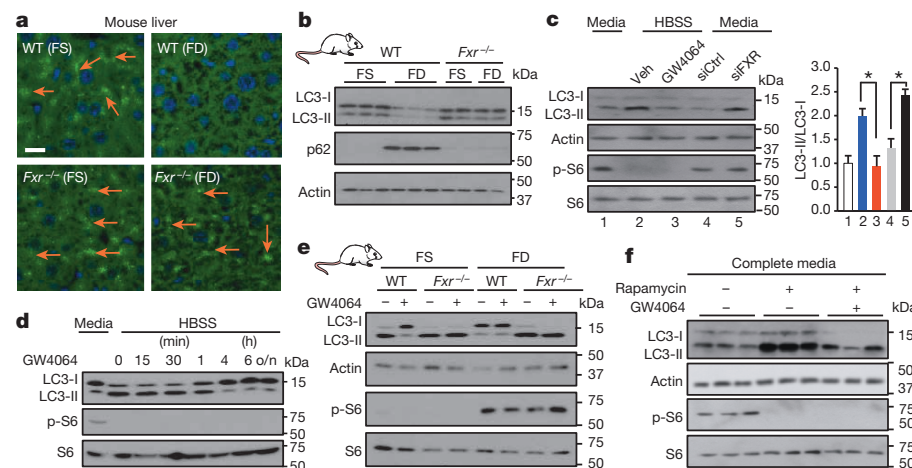


Figure 2 | FXR is a physiological repressor of hepatic autophagy in the fed state.

a, Representative images of GFP-LC3 puncta in livers of fasted or refed mice tail-vein injected with adenovirus containing GFP-LC3 (from a total of five images per mouse; three mice per group for wild type, and two mice per group for *Fxr*^{-/-} mice). Quantification is shown in Extended Data Fig. 2b. **b**, Effects of feeding/fasting on liver LC3 and p62 levels. **c–e**, Effects of GW4064 or downregulation of FXR on LC3 and p-S6 levels in hepatocytes (**c**, **d**) or mice (pooled from three mice) (**e**). **f**, Effects of rapamycin and/or GW4064 on p-S6 and S6 levels in hepatocytes. Data are mean \pm s.e.m., $n = 3$. * $P < 0.05$ (Student's *t*-test).

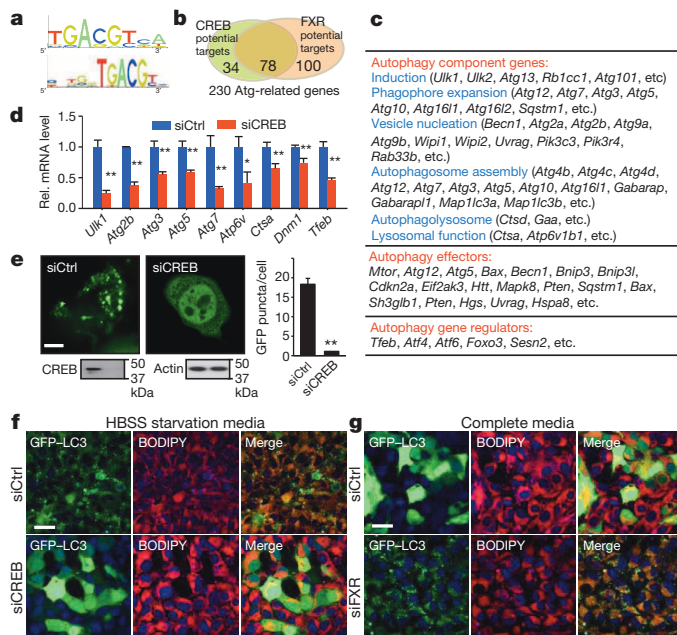


Figure 3 | CREB is a new transcriptional activator of autophagy, and CREB and FXR oppositely regulate lipophagy. **a**, CREB binding motif logos. **b**, Venn diagram showing autophagy genes that contain the FXR and CREB peaks within 10 kb of the TSS. **c**, A list of autophagy genes showing overlapping binding peaks of FXR and CREB. **d**, **e**, Effects of downregulation of CREB on autophagy gene expression (**d**) and GFP-LC3 puncta (**e**) in Hep1c17 cells incubated with HBSS. siCreb, siRNA targeting *Creb*. **f**, **g**, Effects of downregulation of CREB (**f**) or FXR (**g**) on lipophagy in Hep1c17 cells in HBSS starvation media (**f**) or complete media (**g**). Data are mean \pm s.e.m., $n = 3$. * $P < 0.05$, ** $P < 0.01$ (Student's *t*-test).

at CREB-bound autophagy genes, including *Atg7*, *Tfeb* and *Ulk1* in wild-type mice, but not in *Fxr*^{-/-} mice (Fig. 4b and Extended Data Fig. 6c, d). Downregulation of CREB decreased FXR occupancy, suggesting that CREB is important for FXR recruitment (Extended Data Fig. 6e), and

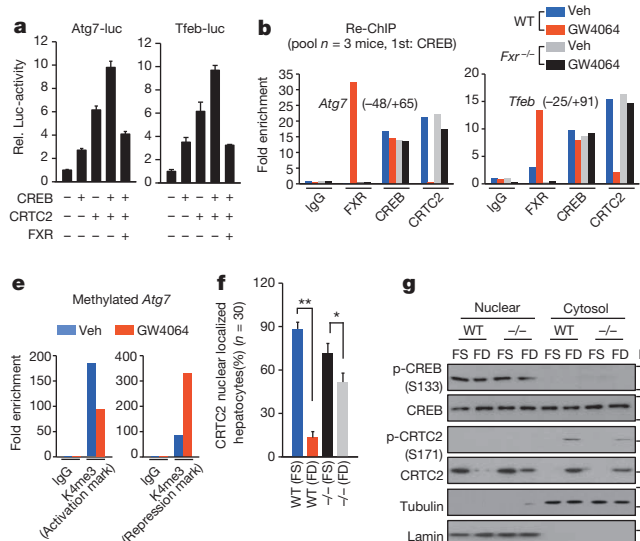


Figure 4 | FXR trans-represses autophagy genes by disrupting the CREB-CRTC2 complex. **a**, *Atg7*-luc and *Tfeb*-luc activities in Hep1c17 cells ($n = 3$, mean \pm s.e.m.). **b**, Re-ChIP: protein occupancy in wild-type and *Fxr*^{-/-} mice. **c**, **d**, Gel-shift assay: an *Atg7* probe contains overlapping CREB and FXR peaks. **e**, ChIP: effects of GW4064 on histone methylations at the *Atg7* promoter. **f**, Effects of feeding/fasting on nuclear localization of CRTC2 in liver ($n = 30$ hepatocytes, mean \pm s.e.m.). * $P < 0.05$, ** $P < 0.01$ (Student's *t*-test). **g**, p-CREB, p-CRTC2 and CRTC2 levels in nuclear and cytoplasmic extracts. **h**, **i**, Re-ChIP: effects of feeding/fasting on protein occupancy in wild-type and *Fxr*^{-/-} mice. **j**, Model: during fasting, autophagy-related genes are induced by CREB-CRTC2, and after feeding, activation of FXR signalling contributes to disruption of the CREB-CRTC2 complex and trans-repression of these genes. BAs, bile acids.

similar effects were observed in mice fed cholic acid chow (Extended Data Fig. 7a, b). In gel-shift assays with an *Atg7* probe, FXR and CRTC2 individually formed complexes with CREB, and FXR caused dissociation of the CREB-CRTC2 complex and formation of a CREB-FXR complex (Fig. 4c). Similar results were observed with *Tfeb* or *Ulk1* probes (Extended Data Fig. 6f–h). Remarkably, activation of FXR was required for the FXR competition (Fig. 4d). Occupancy of FXR at the *Atg7*, *Ulk1* and *Tfeb* genes correlated with increased repressing and decreased activating histone marks and increased occupancy of the corepressors NCoR and SMRT (Fig. 4e and Extended Data Fig. 7c, d), suggesting formation of a repressive transcriptional complex. These results indicate that FXR trans-represses autophagy genes by disrupting the CREB-CRTC2 complex.

We next examined whether physiological activation of FXR after feeding dissociates CRTC2 from the CREB complex. Consistent with previous studies¹⁰, CRTC2 was present in the nucleus in fasted animals and largely excluded from the nucleus in fed wild-type, but not *Fxr*^{-/-}, mice (Fig. 4f and Extended Data Fig. 8a). By contrast, nuclear localization of FXR was markedly increased by feeding or GW4064 treatment, whereas FXR was detected in both the nucleus and cytoplasm during fasting (Extended Data Fig. 8b, c). Fasting increases CREB and CRTC2 activity by phosphorylation and dephosphorylation, respectively¹⁰. p-CREB was detected in the nucleus in fasted mice, whereas p-CRTC2(Ser 171) and CRTC2 were in the cytoplasm of fed mice, but the feeding-induced cytoplasmic localization of p-CRTC2(Ser 171) and CRTC2 was partially reversed in *Fxr*^{-/-} mice (Fig. 4g and Extended Data Fig. 8a). Fasting increased phosphorylated protein kinase A (p-PKA) levels, whereas p-PKB was undetectable and phosphorylated AMP-activated protein kinase (p-AMPK) was not changed (Extended Data Fig. 8d, e). GW4064 inhibition of *Atg7* and *Tfeb* genes was attenuated by expression of a phosphorylation-defective CRTC2(Ser171Ala) mutant constitutively retained in the nucleus¹⁰, but overexpression of FXR reversed this effect (Extended Data Fig. 8f–h). These data suggest that both the phosphorylation status of CREB-CRTC2 and disruption of the CREB-CRTC2 complex by activated FXR are probably important for CREB-dependent fed-state regulation of autophagy genes.

In re-ChIP assays, feeding substantially increased FXR occupancy at CREB-bound genes, whereas occupancy of CRTC2 was decreased

(Fig. 4h, i and Extended Data Fig. 9a, b). By contrast, in *Fxr*^{-/-} mice, the feeding-mediated decrease was reversed, but only partially. Reductions in the mRNA and pre-mRNA levels of these genes after feeding were partially or fully blocked in *Fxr*^{-/-} mice (Extended Data Fig. 9c, d). These partial effects in *Fxr*^{-/-} mice suggest that other meal-related signals also contribute to autophagy gene regulation. Indeed, treatment with insulin, FGF19 or GW4064 downregulated autophagy genes to different extents, in a gene-specific manner (Extended Data Fig. 9e). These findings reveal that disruption of the CREB–CRTC2 complex and trans-repression of autophagy genes by activated FXR is one mechanism by which autophagy is inhibited after feeding.

The present study identifies FXR and CREB as transcriptional regulators of the autophagy gene network (Fig. 4j). During nutrient-deprived fasting, CREB–CRTC2 activates transcription of autophagy-related genes involved in several autophagic processes. After feeding, bile-acid-activated FXR (mimicked by GW4064) interacts with CREB, disrupts the CREB–CRTC2 complex, resulting in decreased nuclear CRTC2 levels, formation of a repressive transcriptional complex, and inhibition of autophagy genes. In addition, feeding-induced changes in the phosphorylation levels of CREB and CRTC2 are also important in disrupting the CREB–CRTC2 complex. The hepatic FXR–CREB axis may also be relevant to other tissues as both proteins are expressed in many tissues. Furthermore, the CREB–CRTC2 complex also activates transcriptional programs of gluconeogenesis and endoplasmic reticulum stress^{10,11}. It will be interesting to see whether FXR more generally inhibits CREB-activated genes.

Notably, both FXR and CREB regulate expression of TFEB, a key activator of autophagy/lysosomal genes^{20,21}. Some autophagy genes seem to be regulated predominantly by either CREB or TFEB or by both (Extended Data Fig. 4d, e and Supplementary Table 5). In addition, other nutrient-sensing regulators, such as peroxisome proliferator-activated receptor- α (PPAR α) (see accompanying manuscript²²), probably transcriptionally regulate autophagy. Only about 45% of FXR autophagy-target genes had shared FXR–CREB binding, so that FXR may act by several mechanisms that are gene- or DNA-context-specific. FXR may trans-repress by disrupting the CREB–CRTC2 complex or by competing with PPAR α for DNA binding, or both mechanisms may occur in the same gene. PKA-activated CREB may also be important early in fasting and dominant after feeding following a short fast, and fatty-acid-activated PPAR α may be more important after prolonged fasting followed by refeeding. Supporting this idea, time-dependent occupancy of CREB–CRTC2 and PPAR α –PGC-1 α after fasting was observed at different regions of the *Atg7* and *LC3a* (also known as *Map1lc3a*) genes (Extended Data Fig. 10).

Autophagy is known to be crucial for cellular survival under extremely stressful conditions, but the present study demonstrates that it occurs physiologically during fasting/feeding cycles by regulation of FXR and CREB, and also mediates lipid catabolic functions. Nutrient-sensing kinases, including mTOR, rapidly decrease activity and/or stability of core autophagic proteins. By contrast, transcriptional regulation by FXR and CREB, which is probably independent of the acutely acting mTOR, is slower and more persistent, which should effectively sustain regulation of autophagy for longer period of times.

Defective autophagy has been implicated in many human diseases, but excess autophagy promotes cell death and also tumour progression^{3,23,24}. In this regard, the nutrient-sensing FXR–CREB axis that tightly regulates the autophagy gene network may present new molecular targets for treating diseases associated with autophagy dysfunction, including metabolic disorders, neurodegenerative disease and cancer.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 October 2013; accepted 8 October 2014.

Published online 12 November 2014.

- Levine, B. & Klionsky, D. J. Development by self-digestion: molecular mechanisms and biological functions of autophagy. *Dev. Cell* **6**, 463–477 (2004).
- Rabinowitz, J. D. & White, E. Autophagy and metabolism. *Science* **330**, 1344–1348 (2010).
- Sachdeva, U. M. & Thompson, C. B. Diurnal rhythms of autophagy: implications for cell biology and human disease. *Autophagy* **4**, 581–589 (2008).
- Mizushima, N. Physiological functions of autophagy. *Curr. Top. Microbiol. Immunol.* **335**, 71–84 (2009).
- Kim, J., Kundu, M., Viollet, B. & Guan, K. L. AMPK and mTOR regulate autophagy through direct phosphorylation of Ulk1. *Nature Cell Biol.* **13**, 132–141 (2011).
- Egan, D. F. et al. Phosphorylation of ULK1 (hATG1) by AMP-activated protein kinase connects energy sensing to mitophagy. *Science* **331**, 456–461 (2011).
- He, C. & Klionsky, D. J. Regulation mechanisms and signaling pathways of autophagy. *Annu. Rev. Genet.* **43**, 67–93 (2009).
- Calkin, A. C. & Tontonoz, P. Transcriptional integration of metabolism by the nuclear sterol-activated receptors LXR and FXR. *Nature Rev. Mol. Cell Biol.* **13**, 213–224 (2012).
- Thomas, C., Pellicciari, R., Pruzanski, M., Auwerx, J. & Schoonjans, K. Targeting bile-acid signalling for metabolic diseases. *Nature Rev. Drug Discov.* **7**, 678–693 (2008).
- Koo, S. H. et al. The CREB coactivator TORC2 is a key regulator of fasting glucose metabolism. *Nature* **437**, 1109–1111 (2005).
- Wang, Y., Vera, L., Fischer, W. H. & Montminy, M. The CREB coactivator CRTC2 links hepatic ER stress and fasting gluconeogenesis. *Nature* **460**, 534–537 (2009).
- Lee, J. et al. Genomic analysis of hepatic farnesoid X receptor (FXR) binding sites reveals altered binding in obesity and direct gene repression by FXR. *Hepatology* **56**, 108–117 (2012).
- Thomas, A. M. et al. Genome-wide tissue-specific farnesoid X receptor binding in mouse liver and intestine. *Hepatology* **51**, 1410–1419 (2010).
- Everett, L. J. et al. Integrative genomic analysis of CREB defines a critical role for transcription factor networks in mediating the fed/fasted switch in liver. *BMC Genomics* **14**, 337 (2013).
- Zhang, X. et al. Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. *Proc. Natl Acad. Sci. USA* **102**, 4459–4464 (2005).
- Singh, R. et al. Autophagy regulates lipid metabolism. *Nature* **458**, 1131–1135 (2009).
- Mizushima, N., Yoshimori, T. & Levine, B. Methods in mammalian autophagy research. *Cell* **140**, 313–326 (2010).
- Inagaki, T. et al. Fibroblast growth factor 15 functions as an enterohepatic signal to regulate bile acid homeostasis. *Cell Metab.* **2**, 217–225 (2005).
- Luo, Q. et al. Mechanism of CREB recognition and coactivation by the CREB-regulated transcriptional coactivator CRTC2. *Proc. Natl Acad. Sci. USA* **109**, 20865–20870 (2012).
- Settembre, C. et al. TFEB links autophagy to lysosomal biogenesis. *Science* **332**, 1429–1433 (2011).
- Settembre, C. et al. TFEB controls cellular lipid metabolism through a starvation-induced autoregulatory loop. *Nature Cell Biol.* **15**, 647–658 (2013).
- Yang, L., Li, P., Fu, S., Calay, E. S. & Hotamisligil, G. S. Defective hepatic autophagy in obesity promotes ER stress and causes insulin resistance. *Cell Metab.* **11**, 467–478 (2010).
- Lee, J. M. et al. Nutrient-sensing nuclear receptors coordinate autophagy. *Nature* <http://dx.doi.org/10.1038/nature13961> (this issue).
- Levine, B. & Kroemer, G. Autophagy in the pathogenesis of disease. *Cell* **132**, 27–42 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Ryerson and D. Tkac for constructing the pGEX4T-FXR clones. We thank M. Montminy for providing Ad-CRTC2 and Ad-CRTC2(S171A). S.-H. Koo for the CREB and CRTC2 plasmids, G. Hotamisligil for Ad-GFP-LC3, and R. Puertollano for the Flag–TFEB plasmid. This study was supported by grants from the National Institutes of Health (DK62777 and DK95842) to J.K.K.

Author Contributions S.S., T.F. and J.K.K. designed research; S.S., T.F., S.-E.C. and S.K. performed experiments; S.S., T.F., S.-E.C., S.K., G.Y., Y.K., B.K. and J.K.K. analysed data; Y.L., R.Z., X.S., W.Z. and J.M. analysed ChIP-seq genomic data, and S.S., T.F., B.K. and J.K.K. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.K.K. (jongsok@illinois.edu).

METHODS

Reagents. Antibodies for FXR (sc-13063), CREB (sc-186), CRTC2 (sc-46272, sc-271912), RNA polymerase II (sc-9001), NCoR (sc-8994), SMRT (sc-1612), p300 (sc-584), PGC-1 α (sc-13067), H3K27me2 (Abcam 050851), H3K4me2 (Abcam 8580), ubiquitin (sc-9133), lamin A (sc-20680), β -tubulin (sc-5274) and actin (sc-1616) were from Santa Cruz Biotechnology. Antibodies for p(S133)-CREB (9198S), LC3 (4108), p62 (5114), S6 kinase (9202), p-S6 kinase (9208) and ATG7 (2631) were from Cell Signaling; and the p(S171)-CRTC2 (bs-3415R) antibody was from Bioss USA. Mouse *Creb1* siRNA was purchased from Thermo Scientific, *Fxr* siRNA and control siRNA from Dharmacon, and *Atg7* siRNA from Life Technologies (AM16708).

In vivo experiments. Eight-to-twelve-week-old male C57BL/6 mice and *Fxr*^{-/-} mice were used for *in vivo* animal studies. Mice were fasted for 10 h and treated with vehicle or GW4064 (30 mg kg⁻¹, intraperitoneally) for 6 h, or were fasted or refed for 6 h. For electron microscopy studies, liver samples were stained with uranyl acetate/lead citrate and then imaged by transmission electron microscopy (EM902A; Carl Zeiss). For FXR activation assays, 10–12-week-old male C57BL/6J mice or *Fxr*^{-/-} mice (Jackson lab) were fasted for 10 h and then injected intraperitoneally with GW4064 (30 mg kg⁻¹) for 6 h or fed normal chow or 0.5% cholic-acid-supplemented chow for 6 h. Livers were then collected for molecular, biochemical and imaging analyses. Adenoviral experiments were performed as previously described^{25–27}. Ad-GFP-LC3 (0.5 \times 10⁹–1 \times 10⁹ active particles per mouse in 200 μ l PBS) was injected via tail vein into wild-type or *Fxr*^{-/-} mice, and 1 week later, the mice were fasted overnight and livers were collected. The frozen or paraffin-embedded liver sections were prepared for staining. *In vivo* experiments were approved by the Institutional Animal Care and Use and Biosafety Committees.

Liver transmission electron microscopy. Mouse liver was fixed using Karnovsky's fixative solution (1% paraformaldehyde, 2% glutaraldehyde, 2 mM CaCl₂ and 100 mM sodium cacodylate, pH 7.4) for 2 h and washed with 100 mM sodium cacodylate, pH 7.4. After post-fixing with 1% osmium tetroxide and 1.5% potassium ferrocyanide for 1 h, mice livers were dehydrated with 50–100% alcohol and stained en bloc with 0.5% uranyl acetate. The liver tissues were embedded in Poly/Bed 812 resin (Pelco), polymerized, and sectioned using a Reichert Jung Ultracut S microtome (Leica) and stained with uranyl acetate and lead citrate. The liver tissue was imaged by transmission electron microscopy (EM902A; Carl Zeiss).

Histology and immunohistochemistry. For *in vivo* detection of hepatic autophagy, frozen or paraffin-embedded samples of liver expressing GFP-LC3 via adenoviral infection were fixed with 4% paraformaldehyde and nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI) and images were taken by confocal microscopy and analysed by ZEN software (Zeiss LSM700). GFP-LC3 puncta were counted in 30 cells for each group and the average number per cell is reported. For monitoring endogenous FXR and CRTC2 localization, paraffin-embedded liver slides were incubated with their primary antibody for 2 h and with secondary antibody, Alexa 647 goat anti-mouse or anti-rabbit IgG for 1 h. Nuclei were labelled with DAPI and images were taken and analysed by confocal microscopy (Zeiss, LSM700). For autophagy studies in cells, Hepa1c1c7 cells were transfected with GFP-LC3 plasmid and/or with 5–20 nM of siRNA for the indicated genes for 24–48 h. Cells were incubated in either complete media or serum free media for 24 h followed by 1 h in HBSS. Cells were treated with GW4064 overnight.

Fluorescent detection of lipophagy. Hepa1c1c7 cells were transfected with siRNAs and plasmids (GFP-LC3, CREB overexpression plasmid). After 36 h, cells were supplemented with 400 μ M oleic acid (O3008, Sigma) for 6 h before incubation in HBSS for 3 h. Cells were fixed with 4% paraformaldehyde (28908, Thermo Scientific) and washed with PBS. A final concentration of 1 μ g ml⁻¹ of BODIPY 581/591 (D-3861, Life Technologies) in 150 mM NaCl was used to stain for lipid for 10 min, and cells were then washed with PBS and counterstained with DAPI. For GW4064 treatment, Hepa1c1c7 cells were transfected with GFP-LC3, and 24 h later, cells were treated with GW4064 or DMSO for 6 h, incubated with HBSS for 3 h, and then stained with BODIPY and counterstained with DAPI.

Biochemical localization study. Mice were fasted or fed overnight. Liver tissues (pooled from three wild-type or *Fxr*^{-/-} mice) were minced and then resuspended in hypotonic buffer and cells were lysed by homogenization. After centrifugation, the nuclear pellet and cytoplasmic supernatant were collected as previously described²⁶. Cytoplasmic β -tubulin and nuclear lamin A were detected as controls for the quality of cellular fractionation.

Hepatocyte isolation and autophagy studies in cells. Primary hepatocytes were isolated by collagenase (0.8 mg ml⁻¹, Sigma-Aldrich Co.) perfusion through the portal vein of mice anesthetized with isoflurane as previously described²⁷. Hepatocytes or Hepa1c1c7 cells were incubated with HBSS starvation media or with complete media and treated with vehicle, GW4064 (100 nM), rapamycin (100 nM), Torin1 (1 μ M), chloroquine (50 μ M) or bafilomycin A₁ (100 nM) for the indicated times, or siRNA (5–20 nM) for 48–72 h.

Liver ChIP and sequential re-ChIP. ChIP and re-ChIP assays were performed as previously described^{25,26}. Livers from wild-type or *Fxr*^{-/-} mice were minced and

incubated in PBS containing 1% formaldehyde for 10 min, and glycine was added to a concentration of 125 mM. Minced liver was resuspended in hypotonic buffer and cells were lysed by homogenization. Nuclei were pelleted, resuspended and sonicated, and chromatin was immunoprecipitated using 1–1.5 μ g of antibody. For re-ChIP assays, CREB was immunoprecipitated first and chromatin was eluted by adding 50 μ l of 10 mM dithiothreitol (DTT) at 37 °C for 30 min. The chromatin samples were then diluted 20-fold with 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM EDTA and 1% Triton X-100, and then re-immunoprecipitated. The immunoprecipitates were extensively washed, chromatin was eluted, and the amounts of genomic DNA were determined by semi-quantitative PCR or quantitative PCR (primer sequences in Supplementary Table 6).

Tfeb-luc and Atg7-luc construction and luciferase assay. DNA fragments near the *Tfeb* gene (chr17:47907235–47908235) and the *Atg7* gene (chr6:114772715–114773373) that contained FXR peaks were amplified by PCR from mouse genomic DNA and cloned into the PGL3-SV40-Luc vector at the Kpn1/Nhe1 and Sac1/BglII sites, respectively, and the cloning was confirmed by sequencing. A FXR mutant was constructed by mutation at the first zinc finger in the DNA binding domain. The values for luciferase activities were normalized to β -gal activities.

GST pull-down, co-immunoprecipitation and qRT-PCR. Sequence encoding the FXR domains, AF1, DNA-binding domain (DBD) or ligand-binding domain (LBD) was amplified by PCR. The PCR product was inserted into the pGEX4T-1 vector at BamHI and XhoI sites. GST-FXR fusion proteins were expressed in *Escherichia coli* BL21/DE3 and purified, CREB was synthesized *in vitro* by *in vitro* transcription and translation (TNT) (Promega, Inc.), and GST pull-down assays were performed. For co-immunoprecipitation experiments, whole liver extracts from mice were prepared and the indicated proteins were immunoprecipitated, and proteins in the immunoprecipitates were detected by immunoblotting. For quantitative reverse transcription PCR (qRT-PCR), total RNA was isolated, complementary DNA was synthesized, and qRT-PCR (primer sequences in Supplementary Table 7) was performed with an Applied Biosystems 7500 qPCR machine. The amount of mRNA was normalized to that of 36B4.

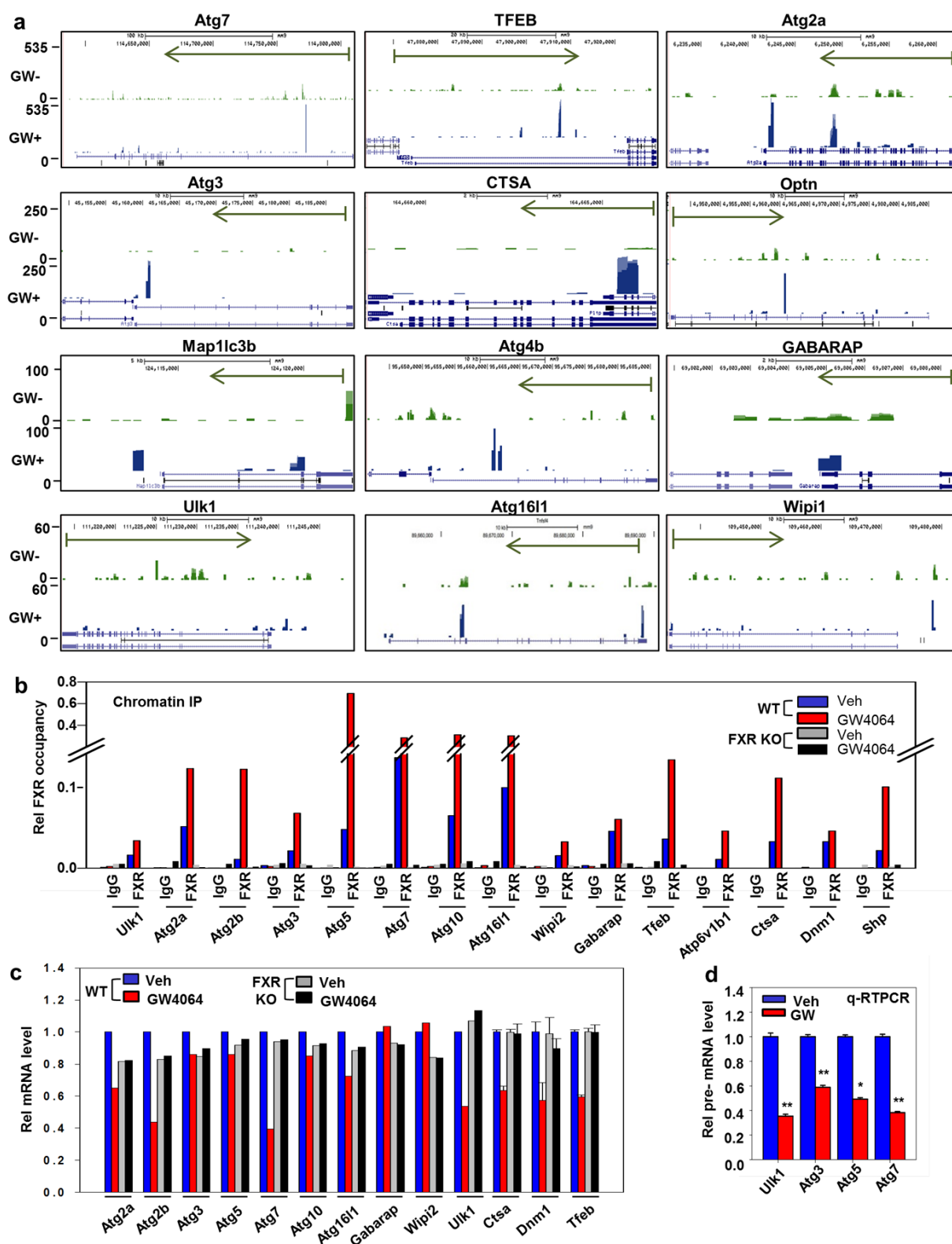
Gel mobility shift assay. Gel-shift assays were done as previously described²⁵. Double-stranded oligonucleotides (31–32 monomeric units) that contain the CREB-binding site from the *Atg7*, *Ulk1* and *Tfeb* genes were labelled with [γ -³²P]ATP and incubated with purified CREB and FXR, and with CRTC2 synthesized by TNT (Promega, Inc.), in the presence of 100 ng poly(deoxyinosinic-deoxycytidylic acid) and 20 μ g BSA. For competition experiments, increasing amounts (1, 3 and 15 ng) of unlabelled oligonucleotide competitors containing a CREB sequence (Supplementary Table 8) from *Atg7*, *Ulk1* and *Tfeb* (a consensus CREB site in intron 1), *Tfeb* (a variant CREB site in the promoter), or *Pepck* as a positive control were incubated 5 min before adding the probe. Binding reactions were analysed by electrophoresis in a low ionic TAE buffer with re-circulation of buffer.

Statistical analyses. Sample sizes were at least $n = 3$ to allow statistical analysis, and for image quantification 10–30 cells were counted. Data are mean \pm s.e.m. P values were generated using Student's t -test ($*P < 0.05$, $**P < 0.01$). No statistical method was used to predetermine sample size.

ChIP-seq data analyses for FXR and CREB. We analysed published ChIP-seq data for FXR^{12,13} and CREB^{15,19}. ChIP-seq reads were uniquely mapped to the mouse reference genome (mm9) by Bowtie with at most two mismatches. MACS (<http://liulab.dfci.harvard.edu/MACS/>) was used to find ChIP-seq peaks with default parameters. Autophagy genes containing FXR or CREB binding peaks within 10 kb of the TSS were considered potential FXR or CREB target genes. Genes were considered to have overlapping FXR–CREB binding peaks if the distance between the midpoints of the FXR and CREB binding peaks was less than 500 base pairs (bp). To identify CREB binding motifs on 78 autophagy genes with overlapping sites, FIMO was used to search for two CREB known motifs within 500 bp of the midpoints between the centres of the overlapping FXR and CREB peaks using position weight matrices, MA0018.1 and MA0018.2, from the JASPAR database with $P < 0.001$.

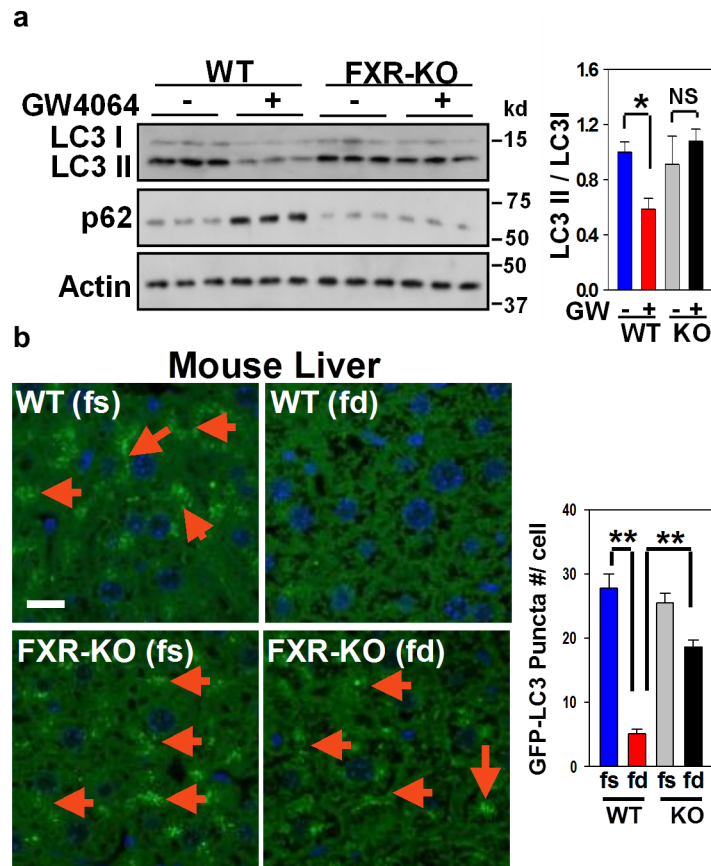
Identification of transcription factor motifs within FXR binding peak regions. For the *de novo* motif analysis within the FXR peaks, the 36 motifs with the least motif score, 1.19, were identified using the Gibbs sampling method, with a maximum motif length set to 30 bp. A total of 50,000 iterations for Markov chain Monte Carlo were used, the top 36 motifs were chosen for transcriptional factor identification. Motif similarity was analysed using the on-line STAMP tools (<http://www.benoslab.pitt.edu/stamp/>).

25. Kemper, J. K. *et al.* FXR acetylation is normally dynamically regulated by p300 and SIRT1 but constitutively elevated in metabolic disease states. *Cell Metab.* **10**, 392–404 (2009).
26. Seok, S. *et al.* Bile acid signal-induced phosphorylation of small heterodimer partner by protein kinase Czeta is critical for epigenomic regulation of liver metabolic genes. *J. Biol. Chem.* **288**, 23252–23263 (2013).
27. Fu, T. *et al.* Aberrantly elevated microRNA-34a in obesity attenuates hepatic responses to FGF19 by targeting a membrane coreceptor β -Klotho. *Proc. Natl Acad. Sci. USA* **109**, 16137–16142 (2012).



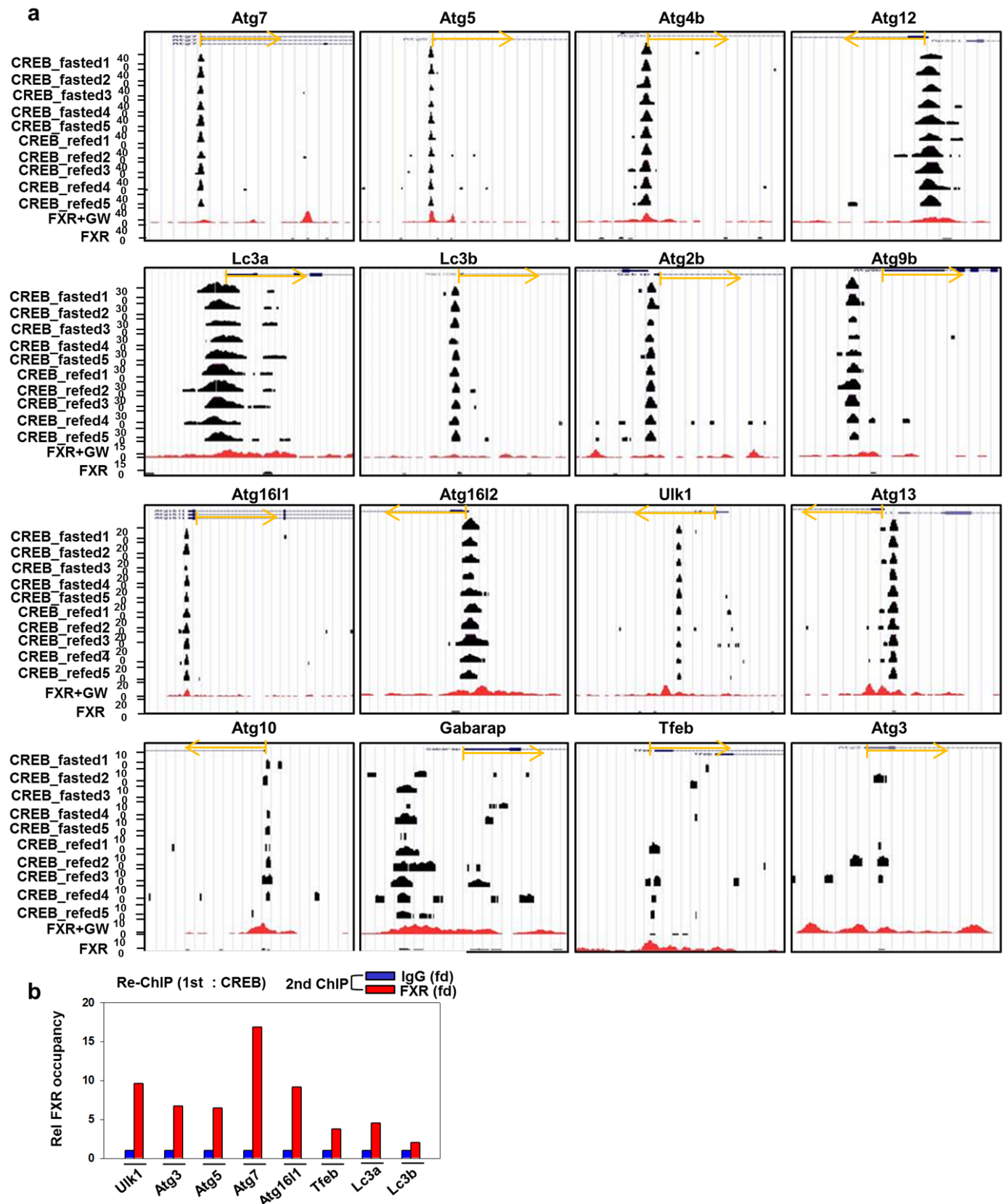
Extended Data Figure 1 | Pharmacological activation of FXR increases its binding to hepatic autophagy genes and decreases their mRNA and pre-mRNA levels. **a**, FXR binding peaks at selected autophagy-related genes. Figures were generated from mouse liver ChIP-seq data using the University of California, Santa Cruz (UCSC) genome browser. Mice were fed a normal chow diet (ND), and treated with vehicle (GW-) or GW4064 (GW+) for 1 h. The direction of gene transcription is indicated by the arrow, and the beginning of the arrow indicates the TSS. **b**, Wild-type or *Fxr*^{-/-} mice were fasted for 10 h and treated with GW4064 (30 mg kg⁻¹, intraperitoneally) or vehicle for 6 h, and liver tissues were then collected. Liver ChIP assays were performed

to confirm occupancy of FXR at the indicated genes. KO, knockout. **c**, Wild-type or *Fxr*^{-/-} mice were fasted for 10 h and treated with GW4064 (30 mg kg⁻¹, intraperitoneally) or vehicle for 6 h. Livers were collected and mRNA levels of the indicated genes were determined by quantitative reverse transcription PCR (qRT-PCR). **d**, Wild-type mice were fasted for 10 h and treated with vehicle or GW4064 (30 mg kg⁻¹, intraperitoneally) for 6 h. Livers were collected and pre-mRNA levels of the indicated genes were determined by qRT-PCR. Data are mean \pm s.e.m., $n = 3$. * $P < 0.05$, ** $P < 0.01$ (Student's *t*-test).



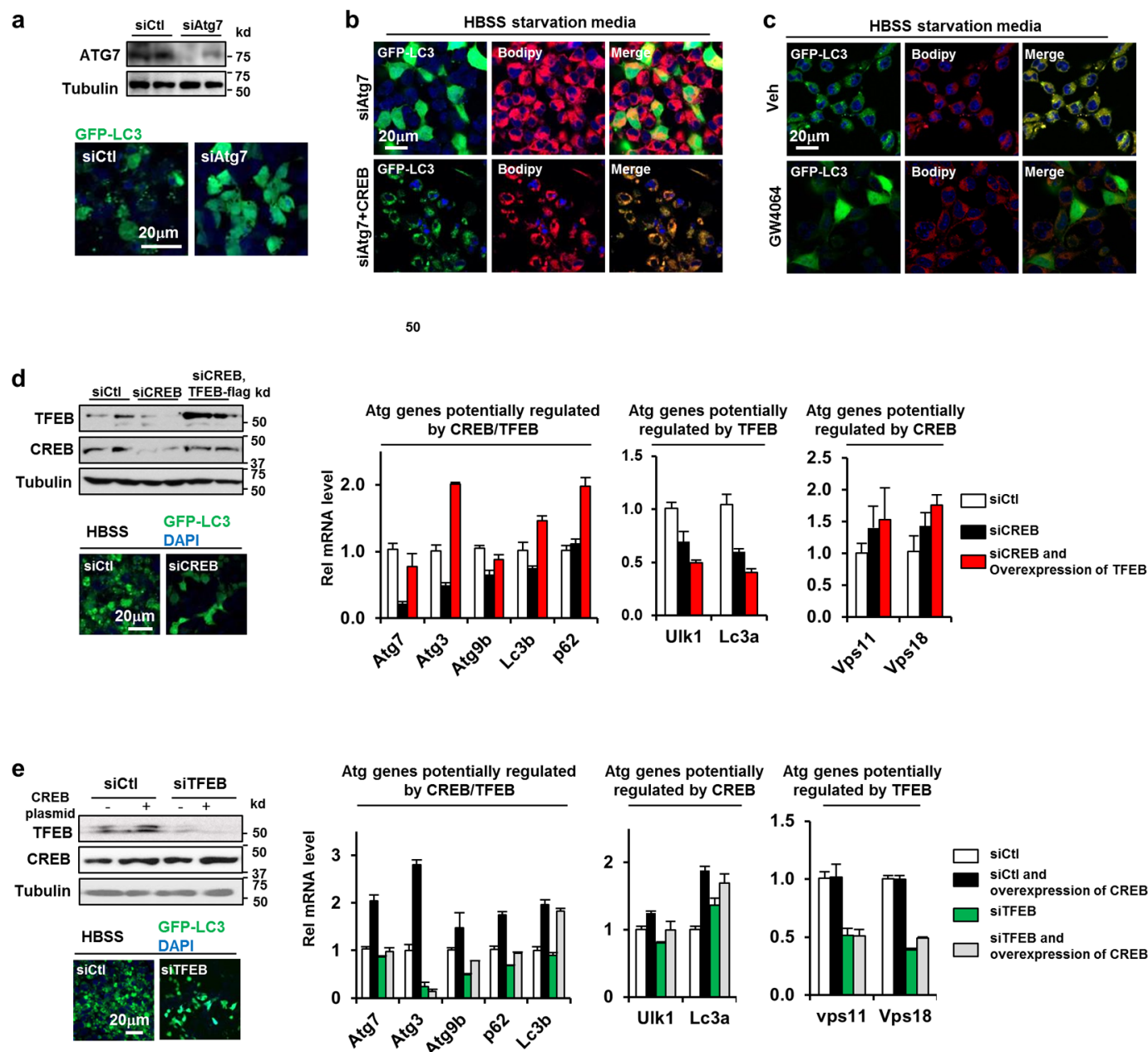
Extended Data Figure 2 | Physiological and pharmacological activation of FXR signalling inhibits hepatic autophagy. **a**, Primary hepatocytes were isolated from wild-type or *Fxr*^{-/-} mice and treated with chloroquine (50 μ M) for 30 min, and then further treated with vehicle or GW4064 (100 nM) for 6 h, and immunoblotting was performed. LC3 and p62 levels in chloroquine-treated hepatocytes and quantification. Representative images of GFP-LC3 puncta

in livers of fasted or refed mice and quantification are shown ($n = 3$). **b**, Mice were tail-vein-injected with Ad-GFP-LC3, and 1 week later, the mice were fasted overnight and livers were collected. The frozen or paraffin-embedded liver sections were prepared for staining. Quantification of GFP-LC3 in each experimental group is shown ($n = 15$ hepatocytes). Data are mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$ (Student's *t*-test).



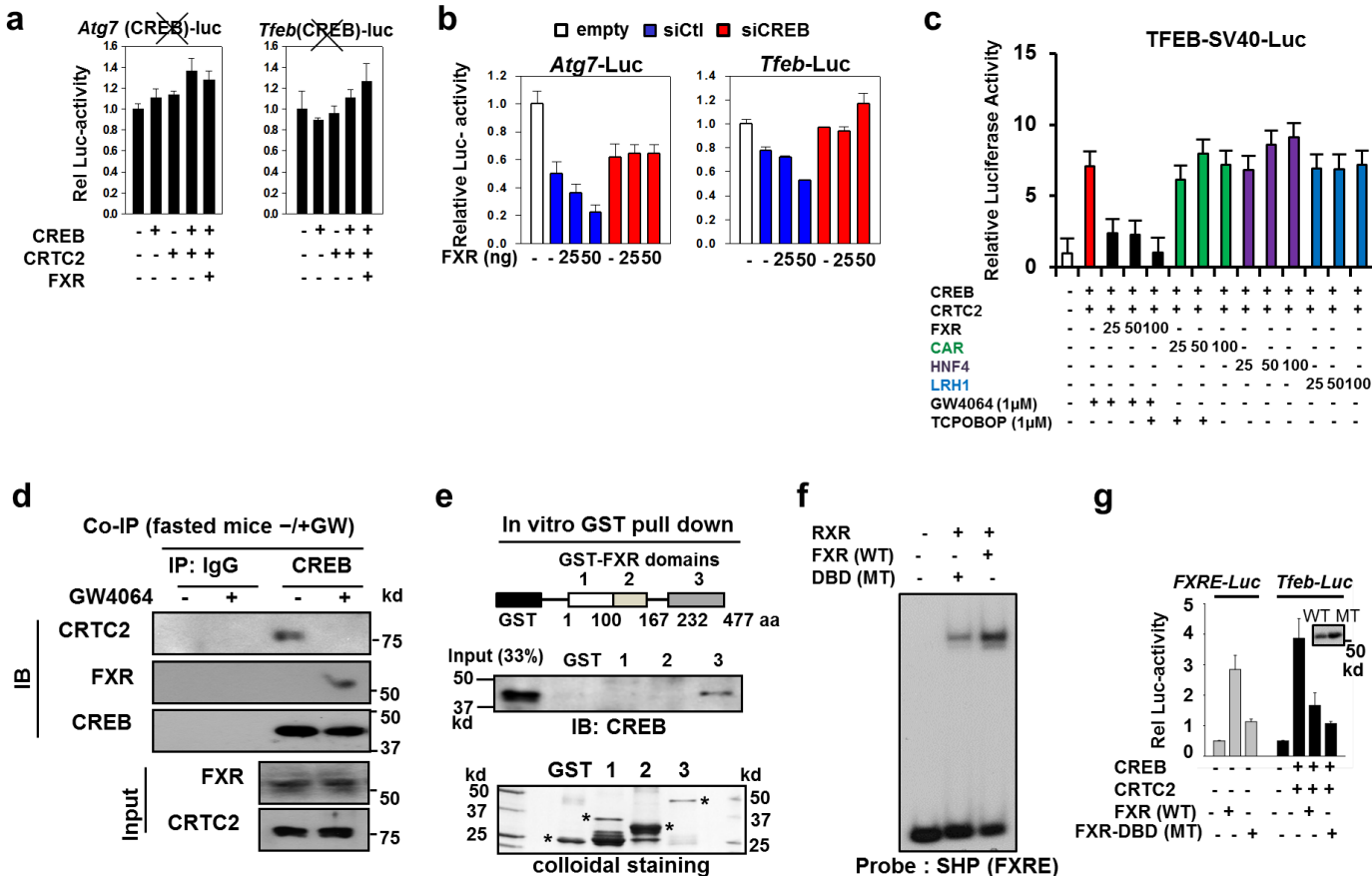
Extended Data Figure 3 | Co-occupancy of FXR with CREB at selected autophagy-related genes in mouse liver. **a**, Figures were generated from liver FXR ChIP-seq data for mice treated with GW4064 (FXR + GW) that are pooled from our group (ref. 12) and from the G. Guo group (ref. 13), and for untreated mice (FXR) from our group. Binding peaks were displayed using the UCSC genome browser. CREB fasted or refed represents CREB1 ChIP-seq

data from ref. 14 for wild-type mice fasted for 24 h or fasted for 24 h and then refed for 2 h, respectively. The direction of gene transcription is indicated by the arrow, and the beginning of the arrow indicates the TSS. **b**, Mouse livers were collected and pooled from two wild-type-fed mice, and sequential re-ChIPs were performed first with CREB antibody and then with FXR antibody or IgG. FXR occupancy at selected autophagy genes was checked by qPCR.



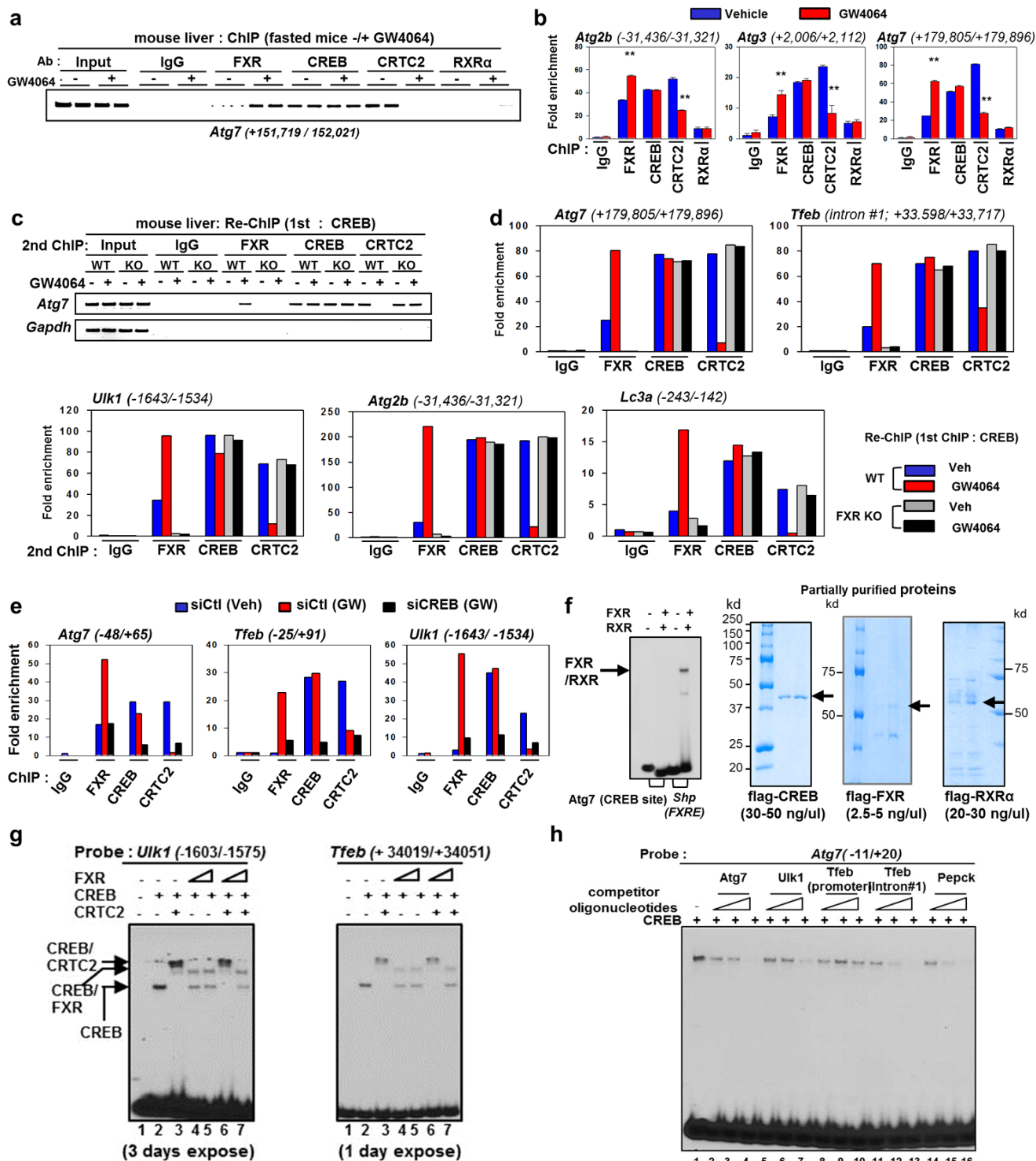
Extended Data Figure 4 | Effects of downregulation of ATG7, CREB1 or TFEB on lipophagy and expression of autophagic and metabolic genes in Hepa1c1c7 cells. **a**, Cells were incubated in HBSS for 3 h to mimic fasting, and ATG7 protein levels were determined by immunoblot. Cells were transfected with the GFP-LC3 plasmid to monitor autophagic puncta in control and autophagy-deficient cells. Images were counterstained with DAPI. **b**, Effects of downregulation of ATG7 and overexpression of CREB on lipophagy. **c**, Effects of treatment with GW4064 on lipophagy in HBSS-incubated starved

Hepa1c1c7 cells. **d**, **e**, Hepa1c1c7 cells were transfected with siRNA or with an expression plasmid as indicated. **d**, TFEB and CREB protein levels were measured by immunoblot. GFP-LC3 fluorescence was used to monitor autophagy in fasted cells. mRNA levels of autophagy genes were measured by qRT-PCR. **e**, TFEB and CREB protein levels were measured by immunoblot. GFP-LC3 fluorescence was used to monitor autophagy in fasted cells. The mRNA levels of autophagy genes were measured by qRT-PCR. Data are mean \pm s.e.m., $n = 3$.



Extended Data Figure 5 | FXR directly interacts with CREB and antagonizes transcriptional activity of CREB. **a**, *Atg7* or *Tfeb* DNA fragments that contained both FXR and CREB peaks sites were cloned and inserted into a luciferase vector. Neither expression of CREB and its coactivator CRTC2 nor expression of FXR increased luciferase activity after mutation of the CREB-binding sites (indicated by X). **b**, Effects of downregulation of CREB in Hepa1c1c7 cells by siRNA on *Atg7* or *Tfeb* luciferase reporter activity. The values for firefly luciferase activities were normalized by β -galactosidase activities. Data are mean \pm s.e.m., $n = 3$. **c**, Hepa1c1c7 cells were transfected with the Tfeb-luc reporter, with CREB and CRTC2 expression plasmids, and with increasing amounts of expression plasmids for FXR, CAR, HNF4 and LXR1. Cells were treated with GW4064 and TCPOBOP for 3 h to activate FXR and CAR, respectively. The values for firefly luciferase activities were

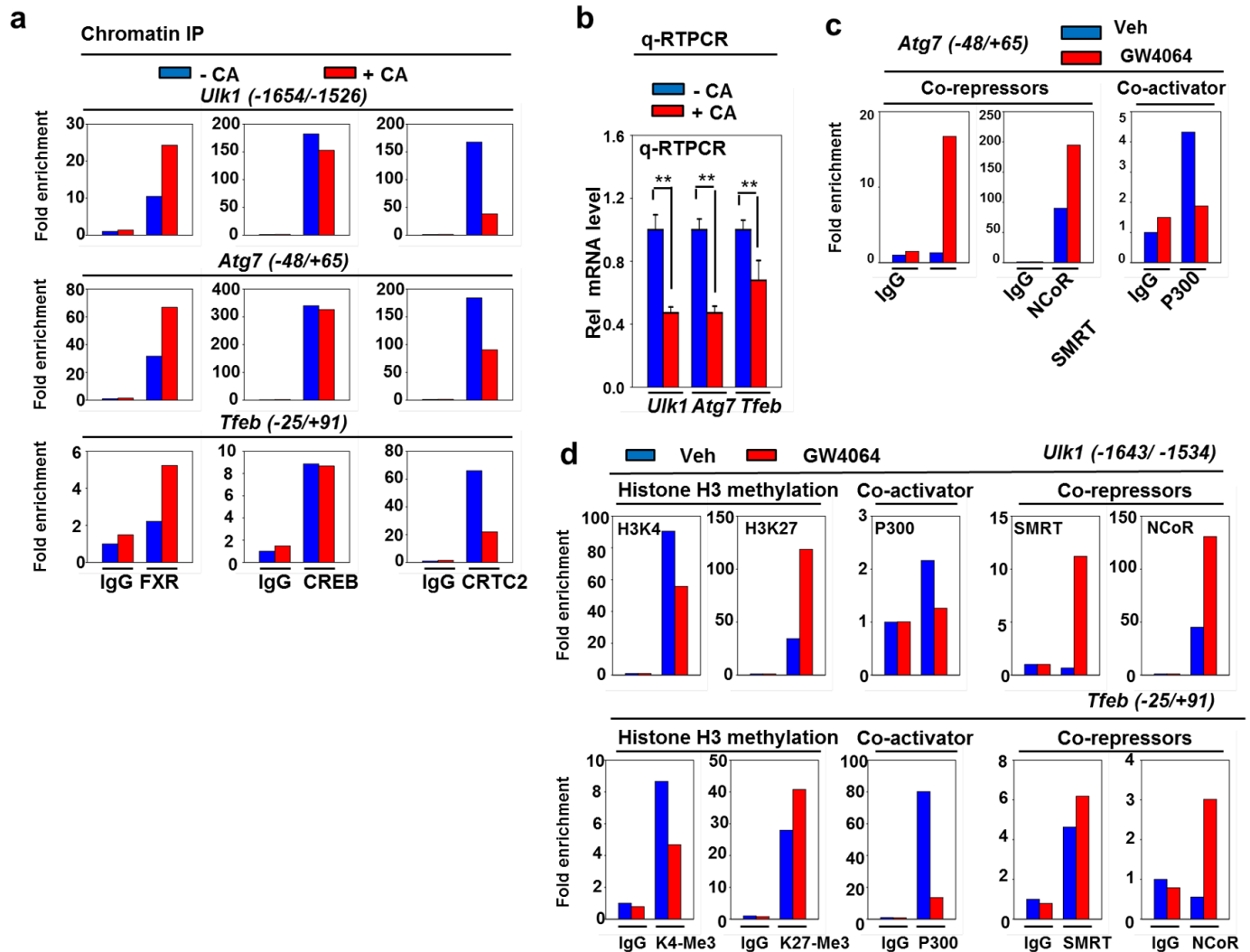
normalized to β -galactosidase activity. **d**, Co-immunoprecipitation assays: mice were fasted for 10 h, and then treated with GW4064 or vehicle for 6 h. Livers were pooled ($n = 3$), CREB was immunoprecipitated from whole-cell extracts, and the levels of CRTC2, FXR and CREB in the immunoprecipitates were determined by immunoblot. **e**, GST pull-down assays: at the top, different fragments of GST–FXR were purified as GST fusion proteins as indicated, and analysed by colloidal staining after SDS–PAGE at the bottom. Binding of CREB to the GST proteins was detected by immunoblot (middle). **f, g**, Effects of DBD mutation on repression of autophagy genes. **f**, *In vitro* gel-shift assays were performed using partially purified RXR and FXR proteins (wild-type or DBD mutant by mutation at the first zinc finger) synthesized by *in vitro* transcription and translation. **g**, Luciferase reporter assays: effects of an FXR DBD mutant (MT) on Teb-luc activity.



Extended Data Figure 6 | Effects of GW4064 treatment on occupancy of FXR, CRTC2 and RXR at CREB-bound autophagy genes in mouse liver.

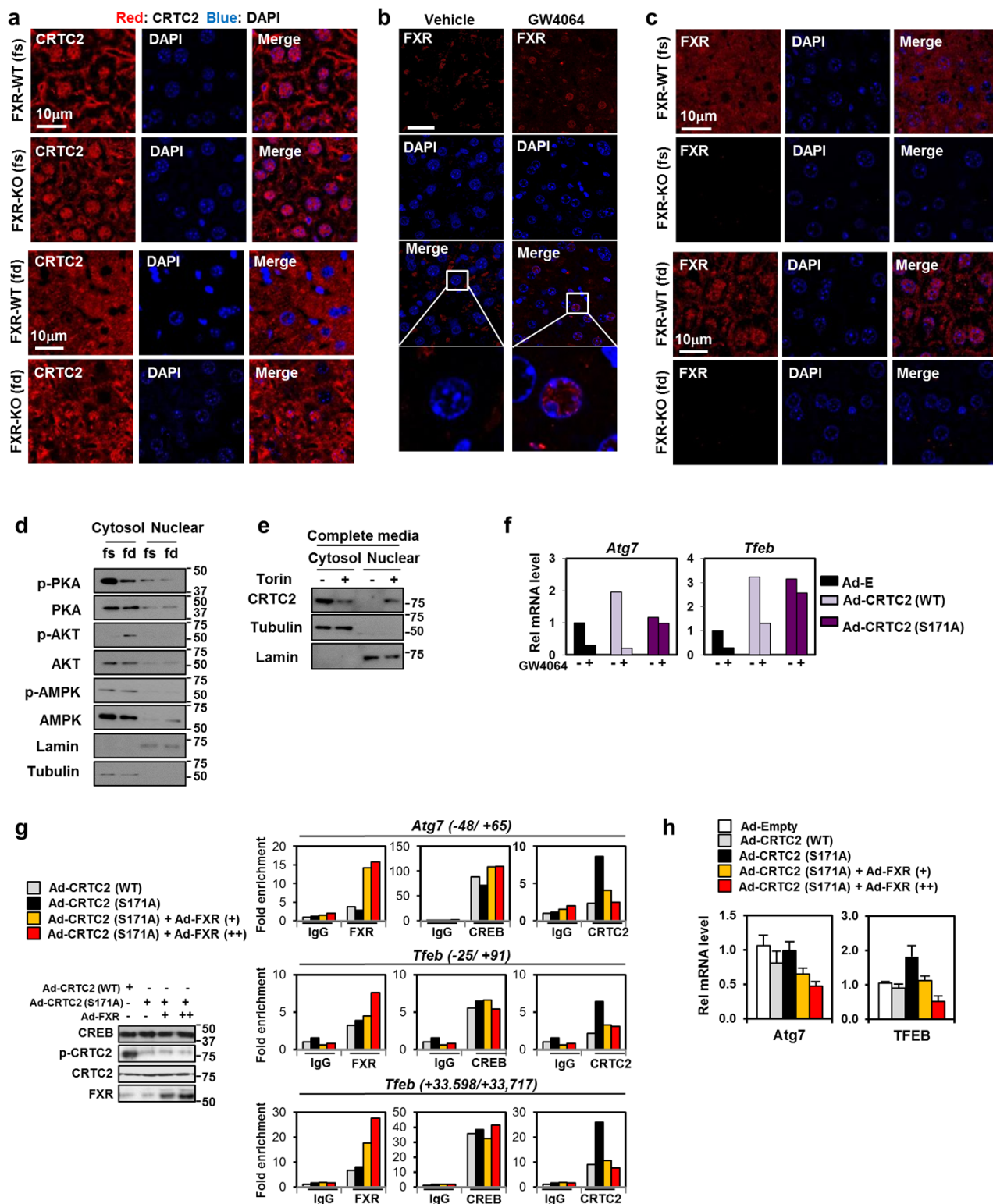
a, b, Liver ChIP: mice ($n = 3$) were fasted for 10 h, and then treated with GW4064 or vehicle for 6 h. Livers were collected and ChIP assays were performed. Occupancy of FXR, CREB, CRTC2 and RXR at the *Atg2b*, *Atg3* and *Atg7* genes was detected by semiquantitative analysis (a) or by qPCR analysis (b). **c, d,** Liver re-ChIP: wild-type or *Fxr*^{-/-} mice were fasted for 10 h and were treated with GW4064 or vehicle for 6 h, livers were pooled from three mice, and re-ChIP assays were performed to detect FXR, CREB and CRTC2 occupancy using semiquantitative PCR analysis (c) or qPCR (d) at the indicated positions of the genes. **e,** Hepa1c1c7 cells were transfected with *Creb* or control siRNA (siCREB or siCtrl) for 36 h, and then treated with GW4064 or vehicle for 6 h. Cells were collected, and ChIP assays were performed.

Occupancy of FXR, CREB and CRTC2 at the *Atg7*, *Tfeb* and *Ulk1* genes was detected by qPCR. **f–h,** Gel-shift assays: FXR, RXR and CREB were partially purified (f, right three panels) and FXR/RXR was incubated with a radiolabelled *Atg7* or *Shp* fragment (f), and FXR, CREB and *in vitro* transcription and translation (TNT)-synthesized CRTC2 was incubated with radiolabelled *Ulk1* or *Tfeb* fragments containing a CREB motif (g). **h,** Competition gel-shift assay: increasing amounts (1, 3 and 15 ng) of oligonucleotide competitors (31–32 mer) containing CREB sites from the indicated genes were incubated with CREB for 5 min before adding the probe. The CREB sequence from the *Pepck* gene serves as a positive control. Similar results were observed from two independent experiments. Data are mean \pm s.e.m., $n = 3$. ** $P < 0.01$ (Student's *t*-test).



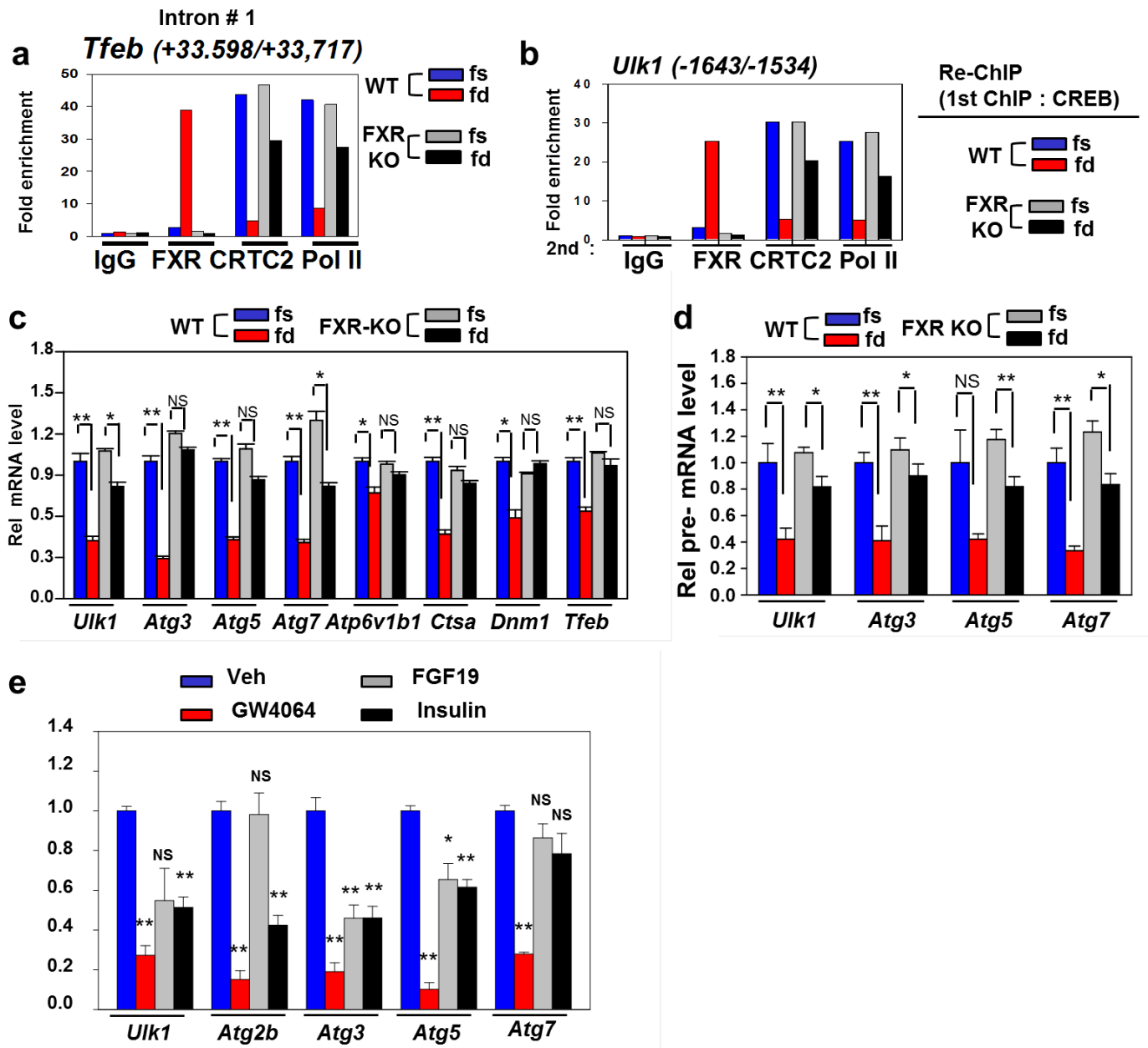
Extended Data Figure 7 | Effect of cholic acid chow on occupancy of FXR, CREB and CRTC2 at selected autophagy genes. Mice fasted for 10 h were refed with cholic acid (CA) chow or normal chow for 6 h, and then livers were collected. **a**, FXR, CREB and CRTC2 occupancies at *Ulk1*, *Atg7* and *Tfeb* genes at the indicated gene positions were determined by qPCR of ChIP assays. **b**, mRNA levels of *Ulk1*, *Atg7* and *Tfeb* genes were measured by qRT-PCR. **c**, **d**, Effects of GW4064 treatment on occupancy of transcriptional coregulators

and histone H3 methylation at *Atg7*, *Tfeb* and *Ulk1* genes in mouse liver. ChIP: mice fasted for 10 h were treated with GW4064 or vehicle for 6 h, livers were collected (pooled from $n = 2$ mice), and ChIP assays were performed. Occupancy of p300, SMRT and NCoR, and levels of histone H3 Lys 4 trimethylation (H3K4me3; gene-activation histone mark) and histone H3 Lys 27 trimethylation (H3K27me3; gene-repression histone mark) at the *Atg7*, *Ulk1* and *Tfeb* promoter regions were detected by qPCR analysis.



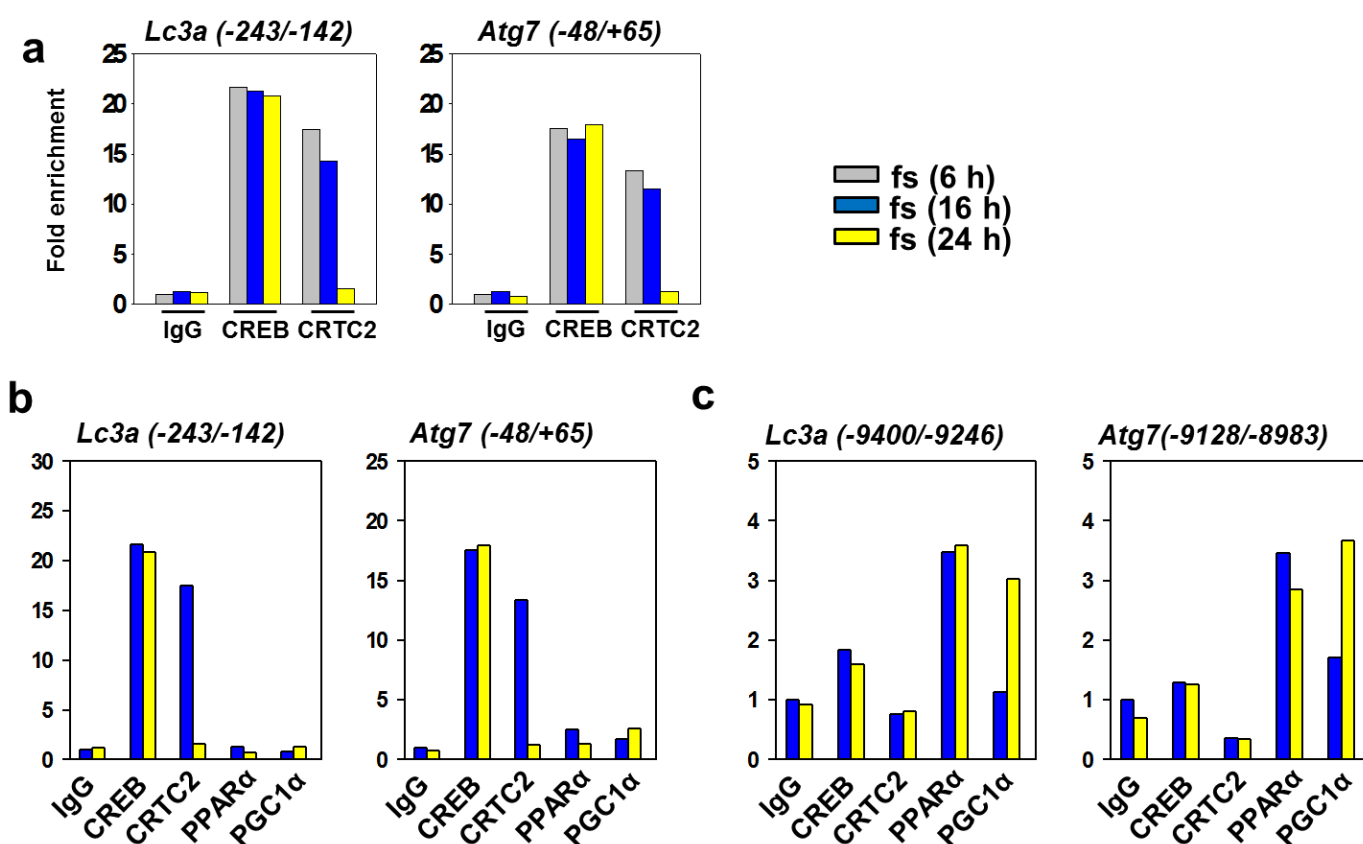
Extended Data Figure 8 | Effects of GW4064 treatment or feeding/fasting on nuclear localization of FXR, CREB and CRTC2, and phosphorylated PKA, PKB and AMPK levels. **a**, Effects of fasting and feeding on localization of CRTC2 in wild-type and *Fxr*^{-/-} mice. CRTC2 staining (red), DAPI (blue), and merged images are shown. **b**, Effects of GW4064 on nuclear localization of FXR in the wild-type mouse liver. **c**, Effects of fasting or feeding on localization of FXR in wild-type and *Fxr*^{-/-} mice. FXR staining (red), DAPI (blue) and merged images are shown. **d**, Nuclear and cytoplasmic extracts were isolated from fasting or fed wild-mice. Lamin and tubulin were measured by immunoblot to monitor the quality of nuclear and cytosol isolation, respectively. Levels of total and phosphorylated PKA, AKT and AMPK were

measured by immunoblot. **e**, Hepa1c7 cells were treated with Torin1 (1 mM) for 1 h and the levels of endogenous CRTC2 in the nucleus and cytosol were determined by immunoblot. **f**, Effects of phosphorylation at Ser 171 of CRTC2 on inhibition of *Atg7* and *Tfeb* genes by GW4064 treatment for 6 h, and then HBSS starvation media to mimic fasting for 1 h in Hepa1c7 cells. **g**, **h**, Hepa1c7 cells were cultured in completed media and infected with Ad-CRTC2 (WT), Ad-CRTC2(S171A) or Ad-FXR as indicated. Expression of indicated proteins was determined by immunoblot (left) and the occupancy of FXR, CREB and CRTC2 was determined by ChIP-qPCR. Consistent results were observed from two independent experiments. Expression of *Atg7* and *Tfeb* genes was detected by qRT-PCR (mean \pm s.e.m., $n = 6$).



Extended Data Figure 9 | Effects of fasting/feeding on occupancy of FXR, CRTC2 and RNA polymerase II at the autophagy-related genes and mRNA and pre-mRNA levels of autophagy genes. **a**, Re-ChIP in wild-type and *Fxr*^{-/-} mice (pooled from three mice). Occupancy of FXR, CREB and CRTC2 at the *Tfeb* gene intron 1 region containing a consensus CREB. **b**, Re-ChIP: effects of GW4064 treatment or feeding on occupancy of FXR, CREB and

CRTC2 at the *Ulk1* gene in wild-type and *Fxr*^{-/-} mice. **c, d**, Wild-type or *Fxr*^{-/-} mice were fasted for 10 h and then fed or fasted for 6 h. mRNA (**c**) and pre-mRNA (**d**) levels of autophagy genes were measured by qRT-PCR. **e**, Primary hepatocytes were treated with GW4064, FGF19 or insulin, and autophagy gene expression was measured by qRT-PCR. Data are mean \pm s.e.m., $n = 3$. * $P < 0.05$, ** $P < 0.01$ (Student's *t*-test).



Extended Data Figure 10 | Fasting time-dependent occupancy of CREB–CRTC2 and PPAR α –PGC1 α at different regions of *Atg7* and *LC3a* genes. **a**, Mice were fasted for 6 h, 16 h or 24 h and ChIP assays were performed. Occupancy of CREB–CRTC2 at the *LC3a* and *Atg7* genes. **b**, **c**, Mice were fasted for 16 h or 24 h and ChIP assays were performed. Occupancy of CREB–CRTC2

and PPAR α –PGC1 α at different FXR binding peak regions of the *LC3a* and *Atg7* genes was detected. A relatively short fasting, 6 or 16 h, seems to be important for CREB–CRTC2 occupancy in the promoter regions of these genes, whereas a longer term 24 h fasting is important for PPAR α –PGC1 α occupancy at distal regions of these genes.

Nutrient-sensing nuclear receptors coordinate autophagy

Jae Man Lee¹, Martin Wagner^{1†}, Rui Xiao¹, Kang Ho Kim¹, Dan Feng^{2†}, Mitchell A. Lazar² & David D. Moore¹

Autophagy is an evolutionarily conserved catabolic process that recycles nutrients upon starvation and maintains cellular energy homeostasis^{1–3}. Its acute regulation by nutrient-sensing signalling pathways is well described, but its longer-term transcriptional regulation is not. The nuclear receptors peroxisome proliferator-activated receptor- α (PPAR α) and farnesoid X receptor (FXR) are activated in the fasted and fed liver, respectively^{4,5}. Here we show that both PPAR α and FXR regulate hepatic autophagy in mice. Pharmacological activation of PPAR α reverses the normal suppression of autophagy in the fed state, inducing autophagic lipid degradation, or lipophagy. This response is lost in PPAR α knockout (*Ppara*^{−/−}, also known as *Nr1c1*^{−/−}) mice, which are partially defective in the induction of autophagy by fasting. Pharmacological activation of the bile acid receptor FXR strongly suppresses the induction of autophagy in the fasting state, and this response is absent in FXR knockout (*Fxr*^{−/−}, also known as *Nr1h4*^{−/−}) mice, which show a partial defect in suppression of hepatic autophagy in the fed state. PPAR α and FXR compete for binding to shared sites in autophagic gene promoters, with opposite transcriptional outputs. These results reveal complementary, interlocking mechanisms for regulation of autophagy by nutrient status.

Overlapping networks govern both acute and longer-term responses to nutrients. In the fasted liver, glucagon induces both rapid glycogen breakdown and transcriptional activation of gluconeogenesis⁶. Nutrient deprivation also acutely regulates nutrient reclamation by autophagy^{7–9}. Among several transcription factors that have been linked to the control of autophagy at the transcriptional level, activation of the basic helix-loop-helix transcription factor EB (TFEB) by fasting is the best described¹⁰.

Nutrient-sensing nuclear receptors are key integrators of metabolic responses. PPAR α is activated by fatty acids in the fasted state, promoting fatty acid oxidation^{4,11,12}. In the fed state, FXR is activated by bile acids returning to the liver along with nutrients, suppressing gluconeogenesis¹³. We proposed that PPAR α and FXR would directly regulate autophagy,

and initially tested the effect of their pharmacological activation in the mouse hepatocyte cell line AML12 expressing a triple fusion protein of monomeric red fluorescent protein (mRFP), green fluorescent protein (GFP) and the autophagosome marker LC3 (mRFP–GFP–LC3)^{14–16}. In autophagosomes, the combination of both RFP and GFP in the triple fusion yields yellow fluorescence, whereas autolysosomal delivery results in red. As expected, the yellow fluorescence in normally cultured cells was converted to red by acute nutrient deprivation (Fig. 1 and Extended Data Fig. 1f). The FXR agonist GW4064 (ref. 17) prevented this process in the nutrient-deprived cells, whereas the PPAR α agonist GW7647 (ref. 18) mimicked it in non-starved cells (Fig. 1a and Extended Data Fig. 1f). Extensive cell-based studies of levels of LC3 conjugated to phosphatidylethanolamine (LC3-II) and other indicators of autophagy confirmed that these responses are dose and time dependent, and indicated that they are not dependent on altered mammalian target of rapamycin complex 1 (mTORC1) activity, but are associated with a decrease in the inhibitory phosphorylation of ULK1 (ref. 19) in response to PPAR α activation, and an increase in response to FXR activation (Fig. 1b, c and Extended Data Fig. 2a–d). Indeed, GW7647 induces autophagic flux even in non-starved cells, whereas GW4064 suppresses it even in starved cells, or cells treated with the mTOR inhibitor Torin1 (ref. 20) (Extended Data Fig. 1a–e). Activation of FXR by bile acids also suppresses LC3-II induction (Extended Data Fig. 2c).

In livers of chow-fed or fasted wild-type C57BL/6J mice, LC3-II levels were increased in the fasted state (Fig. 2a, b and Extended Data Fig. 2e), and messenger RNA expression of *LC3a* (also known as *Map1lc3a*) and other autophagic genes was also induced (Extended Data Figs 2f and 3a–c). PPAR α agonist treatment strongly increased levels of both non-lipidated LC3 (LC3-I) and LC3-II in the fed state in wild-type but not *Ppara*^{−/−} mice²¹, and both were further increased in the fasted state (Fig. 2a and Extended Data Fig. 3b). GW7647 also induced *LC3a* mRNA expression in the fed and fasted states, and these responses were lost in *Ppara*^{−/−} mice (Extended Data Fig. 3a).

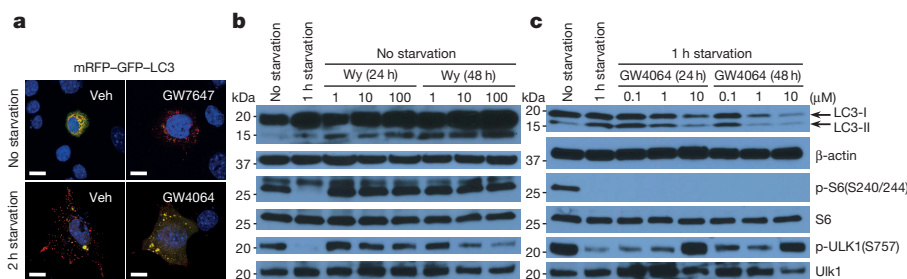


Figure 1 | Activation of PPAR α or FXR controls autophagic flux in mouse hepatocytes. **a**, Representative confocal images (out of 30 cells per condition) of AML12 cells transiently expressing mRFP–GFP–LC3 plasmids followed by treatment of vehicle (Veh), GW7647 (100 nM) or GW4064 (10 μ M) for 24 h. Cells were starved for 2 h (bottom). DNA was counterstained with

4',6-diamidino-2-phenylindole (DAPI) (blue). Scale bars, 20 μ m.

b, c, Immunoblotting of autophagy (LC3-I/II and p-ULK1(Ser757)) or mTORC1 activity (p-S6(Ser240/244)) in AML12 cells treated with indicated doses of Wy-14,643 (Wy) or GW4064 for 24 h or 48 h. GW4064-treated cells were starved for 1 h. kDa, kilodaltons.

¹Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. ²Division of Endocrinology, Diabetes, and Metabolism and the Institute for Diabetes, Obesity, and Metabolism, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19014, USA. [†]Present addresses: Laboratory of Experimental Hepatology, Division of Gastroenterology and Hepatology, Department of Internal Medicine, Medical University of Graz, Auenbruggerplatz 15, A-8036 Graz, Austria (M.W.); Stanford University School of Medicine, Palo Alto, California 94305, USA (D.F.).

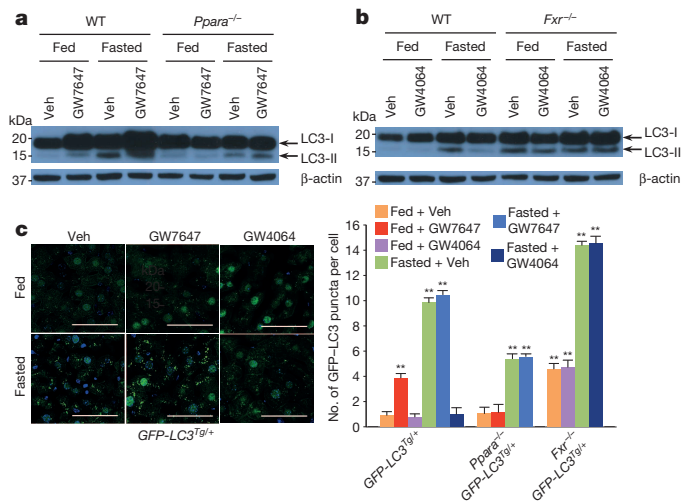


Figure 2 | Activation of PPAR α or FXR controls autophagy in liver. **a, b,** Fed or fasted *Ppara*^{-/-} (**a**), *Fxr*^{-/-} (**b**) or wild-type (WT) mice treated with GW7647 (**a**), GW4064 (**b**) or vehicle. LC3 immunoblots of mouse liver samples. Lanes are pooled ($n = 5$ per group). **c,** Representative confocal images (out of nine tissues sections per condition) of liver GFP-LC3 puncta (green: autophagosomes), counterstained with DAPI (blue: DNA). Fed or fasted GFP-LC3^{Tg/+}, bigenic *Ppara*^{-/-} GFP-LC3^{Tg/+}, or *Fxr*^{-/-} GFP-LC3^{Tg/+} mice treated with vehicle, GW7647 or GW4064. Fixed liver samples analysed by confocal microscopy. Scale bars, 50 μ m. GFP puncta per cell are quantified in the graph. Data represent mean \pm s.e.m. of nine tissue sections ($n = 3$ per group; $**P < 0.01$ versus fed GFP-LC3^{Tg/+} mice treated with vehicle; two-tailed t -test).

In fed or fasted mice treated with the FXR agonist, induction of LC3-II protein in the fasted state was suppressed in wild-type, but not *Fxr*^{-/-} livers (Fig. 2b and Extended Data Fig. 3d). LC3-II levels were unaffected by GW4064 in fed and fasted *Fxr*^{-/-} mice²² (Fig. 2b). At the mRNA level, GW4064 completely suppressed the induction of LC3b in the fasted state (Extended Data Fig. 3c). As expected, GW4064 treatment also suppressed CYP7A1, but induced SHP expression in both the fasted and fed states in wild-type, but not *Fxr*^{-/-} mice (Extended Data Figs 2g and 3c).

These results were confirmed using GFP-LC3 transgenic mice (GFP-LC3^{Tg/+})^{15,16,23} and GFP-LC3^{Tg/+} plus *Ppara*^{-/-} or *Fxr*^{-/-} double mutants. Green puncta indicating autophagosome formation were increased in fasted GFP-LC3^{Tg/+} mice (Fig. 2c). The PPAR α agonist also significantly increased puncta in the fed GFP-LC3^{Tg/+} mice, but not in the *Ppara*^{-/-} GFP-LC3^{Tg/+} double mutants (Fig. 2c and Extended Data Fig. 3e). In the opposite direction, the FXR agonist strongly suppressed induction of puncta in the fasted GFP-LC3^{Tg/+} mice, but not in the fed state, and had no effect in the *Fxr*^{-/-} GFP-LC3^{Tg/+} double mutants (Fig. 2c and Extended Data Fig. 3e). The induction of puncta by fasting was significantly lower in the *Ppara*^{-/-} GFP-LC3^{Tg/+} double mutants,

whereas their number in the fed state was significantly increased in the *Fxr*^{-/-} GFP-LC3^{Tg/+} mutants (Fig. 2c and Extended Data Fig. 3e). These partially defective responses in the double mutants demonstrate that PPAR α is required for the full induction of autophagy by fasting, and FXR is required for its full suppression by feeding.

In agonist- or vehicle-treated fed and fasted livers, quantification of autophagic vesicles by transmission electron microscopy confirmed an increase in response to GW7647 in the fed state, and a strong decrease in response to GW4064 in the fasted state (Fig. 3a). Moreover, fasted livers of *Ppara*^{-/-} mice show compromised formation of autophagic vesicles but increased numbers and size of lipid droplets. Conversely, livers of fed *Fxr*^{-/-} mice show enhanced formation of autophagic vesicles (Extended Data Fig. 4a, c). Autophagosomes induced by the PPAR α agonist frequently contained lipid droplets, suggesting an increase in lipophagy consistent with the role of this receptor in fatty acid oxidation (Extended Data Fig. 4b). This was confirmed in AML12 cells treated with or without oleate to induce lipid droplet formation, and then either starved or not and also treated with vehicle, the PPAR α agonist Wy-14,643 or GW4064. Visualization of LC3 (red) revealed colocalization with lipid droplets labelled with the fluorophore dipyrromethene boron difluoride (BODIPY 493/503; green) in the starved cells, as expected, and also in the non-fasted, Wy-14,643-treated cells (Fig. 3b and Extended Data Fig. 3f). In accord with a functional role for the induction of lipophagy, specific knockdown of either Atg5 or Atg7 significantly blunted the increase in fatty acid oxidation in response to Wy-14,643 in basal and oleate-treated AML12 cells, as indicated by ketone body production (Extended Data Fig. 3g). Similarly, both fasting and GW7647 induced serum β -hydroxybutyrate levels in control homozygous floxed *Atg7* (*Atg7*^{F/F}) mice²⁴, and both responses were decreased in liver specific *Atg7* knockouts (Extended Data Fig. 3h).

Direct transcriptional effects are the most likely explanation for the effect of both PPAR α and FXR on autophagy. Initial results confirmed that total LC3 protein levels are increased and decreased in response to fasting and refeeding, and mRNA expression of *LC3a* and *LC3b* (also known as *Map1lc3b*) and several other autophagic genes is increased in the fasted state and decreased in the refed state (Extended Data Fig. 2e, f). Among a core panel of 63 autophagy-related genes (Supplementary Table 1), 13 were both induced by GW7647 and repressed by GW4064 in wild-type mice, with both responses lost in relevant knockout mice (Fig. 4a and Extended data Fig. 5c), 11 more were responsive only to GW7647, whereas 4 responded only to GW4064 (Extended Data Fig. 5a, b).

To define the basis for PPAR α induction of autophagy-related genes, we determined the mouse liver PPAR α cistrome with or without GW7647. The most significantly enriched binding motif among identified peaks was the known PPAR α -retinoid X receptor (RXR)-binding site (direct repeat 1, or DR1), as expected, and nearly all of the peaks were absent in the *Ppara*^{-/-} cistromes (Extended Data Figs 6a and 7a). PPAR α -binding sites on core autophagy machinery gene loci as well as many regulatory and effector genes were confirmed by standard PPAR α chromatin

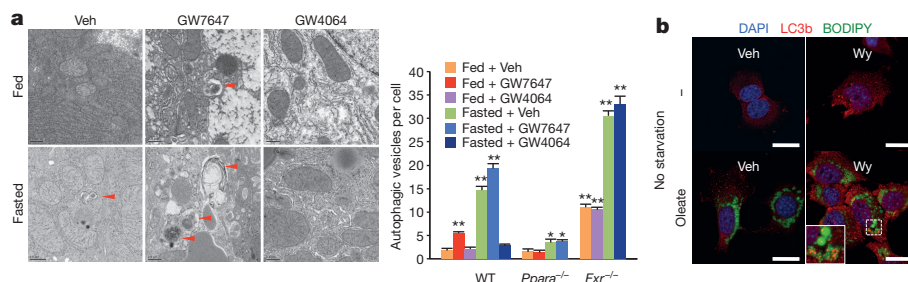


Figure 3 | PPAR α and FXR control autophagic vesicle formation in liver. **a,** Representative mouse liver transmission electron micrographs. Fed or fasted wild-type, *Ppara*^{-/-} or *Fxr*^{-/-} mice treated with vehicle, GW7647 or GW4064. Red arrowheads denote autolysosomes. Scale bars, 0.5 μ m. Autophagic vesicles (autophagosome/autolysosome) per cell are quantified in the graph. Data represent mean \pm s.e.m. for 30 cells per group ($n = 3$ per group;

$*P < 0.05$, $**P < 0.01$ versus vehicle-treated fed wild-type mice; two-tailed t -test). **b,** Co-localization of BODIPY 493/503 (green) with LC3 (red) in AML12 cells treated with vehicle or 10 μ M Wy-14,643 for 24 h, and cultured with or without (–) 125 μ M oleate. DNA stained with DAPI (blue). Scale bars, 20 μ m.

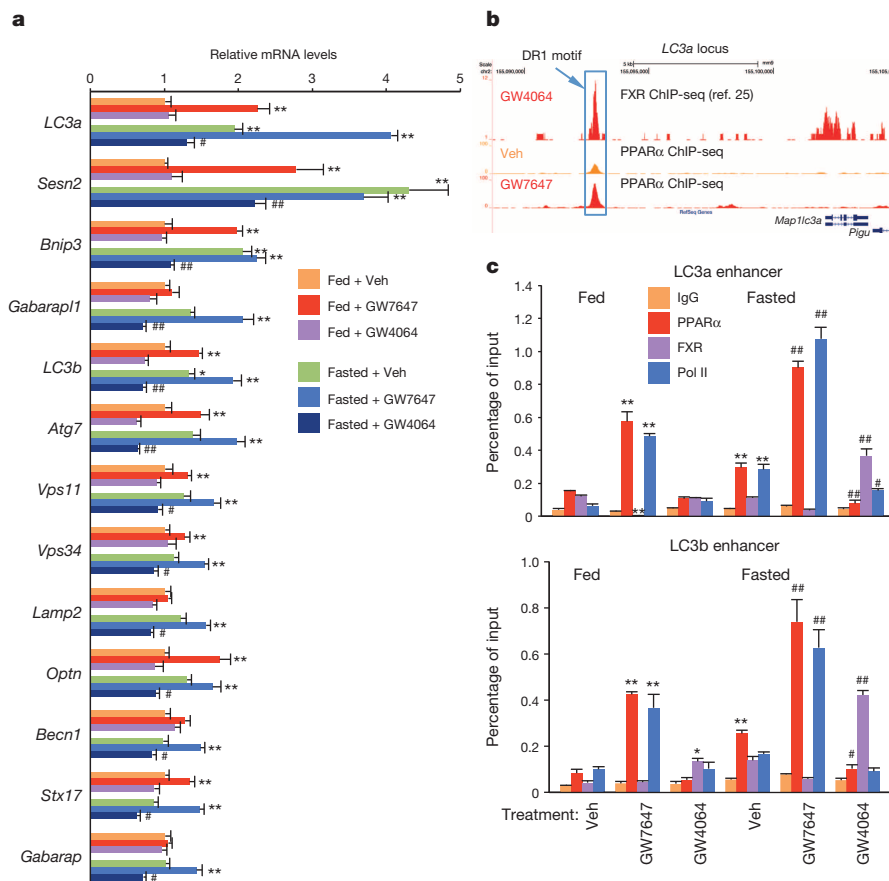


Figure 4 | Transcriptional coordination of hepatic autophagy by nutrient-sensing nuclear receptors in vivo. **a**, Hepatic autophagy-related expression of genes controlled by both PPAR α and FXR in wild-type mouse liver. Fed or fasted mice treated with vehicle, GW7647 or GW4064 ($n = 5$ per group). *Vps34* is also known as *Pik3c3*. **b**, PPAR α and FXR ChIP followed by high-throughput sequencing (ChIP-seq) tracks for LC3a in wild-type mouse

liver. Fed wild-type mice treated with vehicle or GW764 ($n = 4$ per group). Boxed peaks contain DR1 motif. **c**, PPAR α /FXR binding to LC3a/b DR1 region determined by ChIP-qPCR in wild-type mouse liver ($n = 3$ per group). In **a** and **c**, * $P < 0.05$, ** $P < 0.01$ versus vehicle-treated fed wild-type mice; # $P < 0.05$, ## $P < 0.01$ versus vehicle-treated fasted wild-type mice (statistics by two-tailed t -test). Data represent mean \pm s.e.m.

immunoprecipitation (ChIP) and quantitative PCR (qPCR) analysis (Extended Data Figs 6e and 7). Many of the peaks on these genes (Extended Data Fig. 7a), and an additional panel of key autophagy components including *Pink1*, *Optn*, *Vps11* and *Becn1* (Supplementary Fig. 1), were further enhanced by GW7647 treatment, demonstrating that PPAR α agonist treatment not only generates new binding sites, but also further increases PPAR α recruitment to specific target genes (Extended Data Fig. 6b).

The PPAR α cistrome shows that at least 124 of an extended 230 autophagy-related gene list have at least one receptor binding peak within 20 kilobases (kb) from the transcription start site, while a FXR²⁵ cistrome shows 61 potential targets (Extended Data Fig. 6d). A hypergeometric test confirmed that both are highly enriched relative to a random gene pool (Extended Data Fig. 6d). Intriguingly, approximately 27% of agonist-induced PPAR α binding peaks overlapped with 72% of agonist-induced FXR binding peaks (Extended Data Fig. 6c). Consistent with this, peaks for PPAR α and FXR directly overlap at 41 out of the 230 extended autophagic target genes, as exemplified by LC3a (Fig. 4b) and LC3b (Extended Data Fig. 8a). This was unexpected, on the basis of their very different direct (PPAR, DR1) and inverted (FXR, inverted repeat 1 (IR1)) repeat consensus binding sites. However, FXR has also been reported to bind DR1 sites in the *Apoc3* (ref. 26) and *Apoa* (ref. 27) promoters, where it acts as a ligand-dependent transcriptional repressor, raising the prospect that it could repress autophagic gene expression via a similar mechanism.

Quantitative assessment confirmed that GW7646 treatment increased PPAR α binding to peaks containing a single DR1 site in the LC3a

and LC3b enhancers (Fig. 4c) and the *Acox1* promoter (Extended Data Fig. 8a, b), and increased RNA polymerase II recruitment to the genes. Thus, in the fed state, the increased PPAR α binding was able to overcome the relatively silent chromatin state and decrease FXR occupancy for the LC3a and *Acox1* DR1 sites. In the opposite direction, GW4064 did not have much effect on the already silent chromatin in the fed state, in accord with the functional studies (Figs 2 and 3). In the transcriptionally active fasted state, however, the FXR agonist significantly increased FXR binding, decreased RNA polymerase II recruitment, and also decreased PPAR α binding to all three elements (Fig. 4c and Extended Data Fig. 8b). Overall, these results closely match the functional results and indicate that pharmacological activation of each nutrient-sensing nuclear receptor can overcome the chromatin state imposed by the opposite nutrient signal (fed for PPAR α and fasted for FXR). They also indicate that PPAR α and FXR compete for binding to single sites in some chromatin states.

To test this, we focused on the DR1 motifs in the regulatory region of the LC3a and LC3b genes. Cell-based reporter assays with promoter constructs including only the DR1 motifs confirmed the opposite responses, with the PPAR α agonist further inducing, and the FXR agonist suppressing luciferase expression relative to the increased basal transactivation observed with both PPAR α and FXR alone (Extended Data Fig. 8d). In the presence of both receptors, GW4064 suppressed both basal and Wy-14,643-induced expression. Similar results were also observed using a conventional PPAR response element (PPRE) luciferase reporter construct (3 \times PPRE-luc) (Extended Data Fig. 8c). Mutation of the LC3a or LC3b DR1 elements blocked all responses to PPAR α , as expected, but only diminished the responses to FXR (Extended Data Fig. 8d). The basis

for this residual response, which cannot be correlated with direct FXR or FXR–RXR binding *in vitro*, is unclear. The ability of GW4064 to inhibit PPAR α -dependent transactivation of the LC3b DR1 motif is dependent on corepressors, because it was blocked by short interfering RNA (siRNA) knockdown of either SMRT or SHP (Extended Data Fig. 8e). Consistent with these functional results, GW7647 treatment increased p300 co-activator recruitment to the DR1 region of LC3a and LC3b enhancers and increased histone H4 acetylation, an active chromatin mark. By contrast, GW4064 treatment increased binding of the corepressors NCoR and SMRT, and increased levels of the repressive histone H3 trimethyl Lys 27 (H3K27me3) mark at the same DR1 regulatory region, particularly in the fasted state (Extended Data Fig. 9). Overall, we conclude that PPAR α and FXR directly compete for binding to the same DR1 elements at autophagic, and probably additional promoters, with opposite transcriptional outputs.

This mechanism is distinct from, but complementary to, the inhibition of CRTC2 recruitment described in the accompanying manuscript²⁸. Thus, CRTC2 was not significantly recruited to the LC3a DR1 enhancer, as expected, but we confirmed the ability of GW4064 treatment to decrease CRTC2 occupancy to its proximal promoter site in the fasted state (Extended Data Fig. 10a).

These results define new mechanisms for regulation of autophagy by nutrient-sensing nuclear receptors (Extended Data Fig. 10b). Pharmacological activation of either PPAR α or FXR robustly affects hepatic autophagy. GW7467 activates it even in the fed state, which is associated with an increase in lipophagy. GW4064 potentially suppresses the process in the fasted state. The significantly blunted autophagic responses of PPAR α and FXR knockouts to fasting or feeding, respectively, strongly confirm the physiological relevance of the pharmacologically defined responses.

Both nuclear receptors directly induce or repress extensive panels of autophagy-related genes in accord with the overall responses. However, the effect of the synthetic agonists seems greater than the magnitude of the transcriptional responses, suggesting a role for less direct effects. We did not observe effects on mTOR, the dominant upstream regulator of autophagy. However, there were alterations in the inhibitory phosphorylation of ULK1 (Ser 757) that would be consistent with such effects. In addition, induction of the transcriptional repressor SHP is a well-described mechanism for FXR-dependent suppression of genes of bile acid biosynthesis^{29,30}. However, preliminary results indicate that the suppressive effects of GW4064 are maintained in *Shp*^{-/-} (also known as *Nr0b2*^{-/-}) mice (data not shown).

These results reveal intersecting and complementary genomic circuits in which PPAR α and FXR induce and suppress hepatic autophagy, respectively (Extended Data Fig. 10b). They further integrate autophagy, which is often associated with more extreme nutrient stresses, with more physiological nutrient responses. They also establish PPAR α and FXR as potential targets for therapeutic modulation autophagy, which may be useful for liver or other metabolic tissues and could affect the pathogenesis of a wide range of human diseases.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 October 2013; accepted 13 October 2014.

Published online 12 November 2014.

- Mizushima, N., Levine, B., Cuervo, A. M. & Klionsky, D. J. Autophagy fights disease through cellular self-digestion. *Nature* **451**, 1069–1075 (2008).
- Levine, B. & Kroemer, G. Autophagy in the pathogenesis of disease. *Cell* **132**, 27–42 (2008).
- Rabinowitz, J. D. & White, E. Autophagy and metabolism. *Science* **330**, 1344–1348 (2010).
- Evans, R. M., Barish, G. D. & Wang, Y. X. PPARs and the complex journey to obesity. *Nature Med.* **10**, 355–361 (2004).
- Chawla, A., Saez, E. & Evans, R. M. Don't know much bile-ology. *Cell* **103**, 1–4 (2000).

- Nordlie, R. C., Foster, J. D. & Lange, A. J. Regulation of glucose production by the liver. *Annu. Rev. Nutr.* **19**, 379–406 (1999).
- Laplanche, M. & Sabatini, D. M. mTOR signaling in growth control and disease. *Cell* **149**, 274–293 (2012).
- Inoki, K., Kim, J. & Guan, K. L. AMPK and mTOR in cellular energy homeostasis and drug targets. *Annu. Rev. Pharmacol. Toxicol.* **52**, 381–400 (2012).
- Hardie, D. G., Ross, F. A. & Hawley, S. A. AMPK: a nutrient and energy sensor that maintains energy homeostasis. *Nature Rev. Mol. Cell Biol.* **13**, 251–262 (2012).
- Settembre, C. & Ballabio, A. Lysosome: regulator of lipid degradation pathways. *Trends Cell Biol.* <http://dx.doi.org/10.1016/j.tcb.2014.06.006> (21 July 2014).
- Göttlicher, M., Widmark, E., Li, Q. & Gustafsson, J. A. Fatty acids activate a chimera of the clofibrate acid-activated receptor and the glucocorticoid receptor. *Proc. Natl Acad. Sci. USA* **89**, 4653–4657 (1992).
- Keller, H. et al. Fatty acids and retinoids control lipid metabolism through activation of peroxisome proliferator-activated receptor-retinoid X receptor heterodimers. *Proc. Natl Acad. Sci. USA* **90**, 2160–2164 (1993).
- Ma, K., Saha, P. K., Chan, L. & Moore, D. D. Farnesoid X receptor is essential for normal glucose homeostasis. *J. Clin. Invest.* **116**, 1102–1109 (2006).
- Kimura, S., Noda, T. & Yoshimori, T. Dissection of the autophagosome maturation process by a novel reporter protein, tandem fluorescent-tagged LC3. *Autophagy* **3**, 452–460 (2007).
- Mizushima, N., Yoshimori, T. & Levine, B. Methods in mammalian autophagy research. *Cell* **140**, 313–326 (2010).
- Klionsky, D. J. et al. Guidelines for the use and interpretation of assays for monitoring autophagy. *Autophagy* **8**, 445–544 (2012).
- Maloney, P. R. et al. Identification of a chemical tool for the orphan nuclear receptor FXR. *J. Med. Chem.* **43**, 2971–2974 (2000).
- Brown, P. J. et al. Identification of a subtype selective human PPAR α agonist through parallel-array synthesis. *Bioorg. Med. Chem. Lett.* **11**, 1225–1227 (2001).
- Kim, J., Kundu, M., Viollet, B. & Guan, K. L. AMPK and mTOR regulate autophagy through direct phosphorylation of Ulk1. *Nature Cell Biol.* **13**, 132–141 (2011).
- Thoreen, C. C. et al. An ATP-competitive mammalian target of rapamycin inhibitor reveals rapamycin-resistant functions of mTORC1. *J. Biol. Chem.* **284**, 8023–8032 (2009).
- Lee, S. S. et al. Targeted disruption of the α isoform of the peroxisome proliferator-activated receptor gene in mice results in abolishment of the pleiotropic effects of peroxisome proliferators. *Mol. Cell. Biol.* **15**, 3012–3022 (1995).
- Sinal, C. J. et al. Targeted disruption of the nuclear receptor FXR/BAR impairs bile acid and lipid homeostasis. *Cell* **102**, 731–744 (2000).
- Mizushima, N., Yamamoto, A., Matsui, M., Yoshimori, T. & Ohsumi, Y. In vivo analysis of autophagy in response to nutrient starvation using transgenic mice expressing a fluorescent autophagosome marker. *Mol. Biol. Cell* **15**, 1101–1111 (2004).
- Komatsu, M. et al. Impairment of starvation-induced and constitutive autophagy in *Atg7*-deficient mice. *J. Cell Biol.* **169**, 425–434 (2005).
- Thomas, A. M. et al. Genome-wide tissue-specific farnesoid X receptor binding in mouse liver and intestine. *Hepatology* **51**, 1410–1419 (2010).
- Claudel, T. et al. Farnesoid X receptor agonists suppress hepatic apolipoprotein CIII expression. *Gastroenterology* **125**, 544–555 (2003).
- Chennamsetty, I. et al. Farnesoid X receptor represses hepatic human APOA gene expression. *J. Clin. Invest.* **121**, 3724–3734 (2011).
- Seok, S. et al. Transcriptional regulation of autophagy by an FXR–CREB axis. *Nature* <http://dx.doi.org/10.1038/nature13949> (this issue).
- Kerr, T. A. et al. Loss of nuclear receptor SHP impairs but does not eliminate negative feedback regulation of bile acid synthesis. *Dev. Cell* **2**, 713–720 (2002).
- Wang, L. et al. Redundant pathways for negative feedback regulation of bile acid production. *Dev. Cell* **2**, 721–731 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank N. Mizushima for the GFP-LC3^{Tg/+} mice; T. Yoshimori for the mRFP–GFP–LC3 plasmid; M. Komatsu for the *Atg7*^{+/F} mice; D. Townley and M. Mancini for transmission electron microscopy and confocal microscopy; the members of the Moore laboratory for comments and additional support. Core facilities supported by grants U54 HD-07495-39, P30 DX56338-05A2, P39 CA125123-04 and S10RR027783-01A1. Next-generation sequencing was performed by the Functional Genomics Core of the Penn Diabetes Research Center (DK19525). This work was supported by funding from the Alkek Foundation and the Robert R. P. Doherty Jr-Welch Chair in Science to D.D.M., and R01 DK49780 and DK43806 to M.A.L.

Author Contributions J.M.L. conceived the project, designed and performed most experiments, interpreted results, and co-wrote the manuscript. M.W. performed animal experiments and participated in discussion of the results. R.X. analysed PPAR α and FXR ChIP-seq data, and designed primers for PPAR α ChIP-qPCR. K.H.K. performed ChIP assays and molecular cloning. D.F. performed PPAR α ChIP-seq. M.A.L. supervised experimental designs. D.D.M. conceived the project, supervised experimental designs, interpreted results, and co-wrote the manuscript.

Author Information PPAR α ChIP-seq data sets have been deposited in the NCBI Gene Expression Omnibus with the accession number GSE61817. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.D.M. (moore@bcm.edu).

METHODS

Materials. C57BL/6J and *Ppara*^{-/-} mice were from the Jackson Laboratory; *Fxr*^{-/-} and *GFP-LC3* transgenic mice were previously described²³. Antibodies to LC3 (NB600-1384) were from Novus Biologicals; rabbit antibodies to phosphorylated Ser 240/244 S6 (5364), S6 (2217), phosphorylated Thr 37/46-4E-BP1 (2855), 4E-BP1 (9452), anti-mouse IgG-horseradish peroxidase (HRP) (7076) and β -actin (5125) were from Cell Signaling Technology; antibodies to normal rabbit IgG (sc-2027), PPAR α (sc-9000x), FXR (sc-13063x), Pol II (sc-899x), p300 (sc-585x), CRTC2/TORC2 (sc-6714x) and anti-rabbit IgG-HRP (sc-2370) were from Santa Cruz Biotechnology; antibody to histone H3K27me3 (39155) was from ACTIVE MOTIF. GW7647 and Wy-14,643 were from Cayman chemicals; GW4064 was from Tocris bioscience; bafilomycin A₁ was from Enzo Life Science; AML12 cells were from ATCC (CRL-2254). GFP antibody (11814460001), PhosSTOP (04906837001) and complete protease cocktail (11836170001) were from Roche; insulin-transferrin-selenium (ITS) (41400-045), lipofectamine 2000 (11668019), BODIPY 493/503 (D-3922), Alexa Fluor 594 goat anti-rabbit IgG (A11037), Stealth siRNAs-NCoR (NcoR1, MSS208758), SMRT (NcoR2, MSS20912), SHP (NR0b2, MSS239500) and HIGH_GC were from Invitrogen. The mRFP-GFP-LC3(ptfLC3) plasmid was a gift from T. Yoshimori (Addgene plasmid 21074). Anti-rabbit p62/SQSTM1 (P0067), anti-rabbit LC3B (L7543), oleic-acid-albumin (O3008-5ML) and dexamethasone (D4902-25MG) were from Sigma. Protein A Sepharose CL-48 (17-0780-01) was from GE Healthcare; VECTASHIELD (H-1200) was from Vector Laboratories; SacI-HF (R3156S) and BglII (R0144S) were from New England Biolabs. Pierce BCA Protein Assay Kit (23227) was from Thermo Scientific; SMRT antibody (ab24551) and β -hydroxybutyrate assay kit (ab83390) were from abcam. Anti-acetyl-histone H4 antibody (06-866) was from EMD Millipore. Polyclonal rabbit antibody against NCoR was as described³³. Other reagent information not shown here is described in the relevant methods and references.

Animal studies. All animal studies and procedures were approved by the institutional Animal Care and Use Committee of the Baylor College of Medicine. Male *GFP-LC3*^{Tg/+} mice were bred with either female *Ppara*^{-/-} or *Fxr*^{-/-} mice to generate bigenic *Ppara*^{-/-} *GFP-LC3*^{Tg/+} or *Fxr*^{-/-} *GFP-LC3*^{Tg/+} mice. *Ppara*^{-/-} and *Fxr*^{-/-} mice were pure C57BL/6 background. C57BL/6J mice were wild-type controls. Eight-to-ten-week-old male wild-type C57BL/6J, *Ppara*^{-/-}, *Fxr*^{-/-}, *GFP-LC3*^{Tg/+}, *Ppara*^{-/-} *GFP-LC3*^{Tg/+} or *Fxr*^{-/-} *GFP-LC3*^{Tg/+} mice were orally gavaged with vehicle (dimethylsulphoxide (DMSO) in 4:1 of PEG-400 and Tween 80), GW7647 (5 mg kg⁻¹ body weight) or GW4064 (100 mg kg⁻¹ body weight) twice a day (first injection: 00:00, second injection: 12:00). Mice were fed *ad libitum* or fasted for 24 h from 18:00 (day 1) to 18:00 (day 2) and then sacrificed to collect blood and livers. Collected liver tissues were immediately frozen in liquid nitrogen for molecular studies. The refed condition was 24 h fasted mice followed by feeding those mice with normal chow diets for another 24 h. To avoid circadian issues, all mice were sacrificed at 17:00–18:00. *GFP-LC3*^{Tg/+} mice were a gift from N. Mizushima.

Primary hepatocyte culture. Eight-to-ten-week-old male *GFP-LC3*^{Tg/+} mice were anesthetized with Avertin (Sigma Aldrich, T48402, 240 mg kg⁻¹ body weight) by intraperitoneal (i.p.) injection. The liver perfusions were done by injecting butterfly needle into the portal vein and providing the following solutions sequentially: 50 ml HBSS (ATLANTA biologicals, B31350) supplemented with 10 mM HEPES (Invitrogen, 24020-117) and 0.5 mM EGTA and then 50 ml of HBSS supplemented with 100 U ml⁻¹ of collagenase IV (Invitrogen, 17101-015), and 0.05 mg ml⁻¹ of trypsin inhibitor (Invitrogen, 17075-029). The perfused liver was carefully taken out, put onto a Petri dish with 25 ml of hepatocyte wash media added (Invitrogen, 17704-024), and massaged with two cell scrapers until the liver came apart with only connective tissue left behind. Dissociated cells were passed through a mesh funnel into a 50 ml falcon tube. After centrifugation at 600 r.p.m. for 2 min, the cell pellet was resuspended in hepatocyte wash media consisting of carefully overlaid Percoll solution (final 40%) pre-mixed with 150 mM NaCl. After centrifugation at 600 r.p.m. for 10 min, the collected cell pellet was washed twice with hepatocyte wash media, and then suspended in Williams' E medium (Invitrogen, 12551-032) supplemented with 10% FBS, 1% penicillin/streptomycin, ITS and glutaMAX (Invitrogen, 35050-061). All drug treatments or starvation experiments were performed 4–6 h after the initial plating on the culture dishes.

Cell culture and transfection. AML12 cells were maintained in the following media: (normal) DMEM/F12 high glucose (Invitrogen, 11330-057) supplemented with 10% FBS, 1% ITS, 1% penicillin/streptomycin antibiotics and 40 ng ml⁻¹ dexamethasone; (starvation) HBSS media with Ca²⁺ and Mg²⁺ supplemented with 10 mM HEPES (Invitrogen, 15630). mRFP-GFP-LC3 plasmids were transfected into AML12 cells grown on Lab-tek II Chamber Slide (Nunc, 154461) with Lipofectamine 2000 according to the manufacturer's protocol for 24 h followed by drug treatments for 24 h. Transfected cells were starved for 2 h, and then washed three times with cold PBS, fixed with 3% paraformaldehyde (PFA) in PBS for 10 min at room temperature, and washed three times with PBS. Cells were counterstained with mounting solution containing DAPI. All drugs were dissolved in DMSO. Vehicle: complete media

containing 0.1% DMSO. Starvation: HBSS media with 10 mM HEPES. Autophagic fluxes were quantified as follows: autolysosomes = RFP-positive vesicles – GFP-positive vesicles. Numbers of autolysosomes induced by 2 h starvation plus vehicle were set as 100%.

BODIPY and immunofluorescence assay in cells. AML12 cells were plated on two-chamber Lab-tek II dishes at 30–50% confluent cells per chamber. The next day, cells were treated with vehicle (0.1% DMSO), 10 μ M Wy-14,643 or 1 μ M GW4064 in complete medium with or without 125 μ M oleate for 24 h. Vehicle or GW4064-treated cells were starved for 2 h in HBSS medium before fixation of cells. Cells were washed twice with PBS followed by fixation with 3% PFA for 20 min at room temperature. After fixation, cells were washed four times with PBS followed by incubation with blocking buffer (1.5 g glycine, 3 g BSA and 2 ml 0.5% (w/v) saponin in 100 ml PBS) for 45 min. Cells were incubated with rabbit anti-LC3b antibody (1:400 dilution in antibody diluent: 100 mg BSA, 2 ml 0.5% saponin in 100 ml PBS) overnight at 4 °C. Cells were washed four times (10 min each) with PBS followed by incubation with the secondary antibody (Alexa Fluor 594 goat anti-rabbit IgG, 1:200 dilution in antibody diluent) and BODIPY 493/503 (1 μ g ml⁻¹ concentration) for 1 h at room temperature. After that, cells were washed four times with PBS followed by DAPI staining with mounting solution. Cell images were obtained using a confocal microscope (NIKON A1Rsi dual scanner). Orange and yellow dots were defined as a lipophagic vacuole, which was quantified in at least 30 cells per condition using ImageJ software.

RNA purification and qPCR analysis. Total RNA was isolated from snap-frozen liver tissues using Trizol Reagent (Invitrogen, 15596-026) and prepared for the complementary DNA using QuantiTect reverse transcription kit (Qiagen, 205311). Hepatic gene expression was determined by qPCR using FastStart SYBR Green master (ROX) (Roche, 04673484001). mRNA levels were normalized by the *36B4* (also known as *Rplp0*) gene. qPCR primer information for autophagy-related genes is listed in Supplementary Table 1, and all other primers were purchased from Qiagen.

Immunoblot analysis. Cells cultured in 10-cm dishes were solubilized in 150 μ l of RIPA buffer (25 mM Tris-HCl, pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS and 1 mM EDTA) supplemented with protease and phosphatase inhibitors. Liver tissues (50 mg) were homogenized in 1 ml RIPA buffer supplemented with protease and phosphatase inhibitors, followed by brief sonication. Protein concentration was determined using Pierce BCA Protein Assay Kit. Total proteins (25–50 μ g) were loaded on either 12% or 4–20% Mini-PROTEAN TGX Precast Gel (Bio-rad, 456-1043 or 456-1093), transferred onto Immuno-Blot PVDF (polyvinylidene difluoride) membrane (Bio-rad, 162-0177) and analysed by immunoblot analysis using the ECL solution (Thermo Scientific, 34086, or Millipore, WBKLS0500).

Histology and immunofluorescence. Liver tissues were collected and fixed in 4% PFA in PBS overnight at 4 °C. The tissues were treated with 15% sucrose in PBS at 4 °C for at least 4 h and then with 30% sucrose in PBS at 4 °C for overnight. Tissue samples were embedded in Tissue-Tek OCT (Sakura Finetek, 4583) and stored at –70 °C. The samples were sectioned at 5–7 μ m thickness, air-dried for 30 min, and stored at –20 °C until use (Comparative Pathology Laboratory at Baylor College of Medicine). Images were taken on a confocal microscope (NIKON A1Rsi dual scanner). GFP-LC3 puncta were quantified in three independent visual fields from at least three independent mice using ImageJ software.

ChIP assays. ChIP from mouse livers was performed as described previously³⁴. For PPAR α ChIP-seq and ChIP-qPCR (Extended Data Fig. 7), livers were collected from fed wild-type or *Ppara*^{-/-} mice treated with vehicle or GW7647 (5 mg kg⁻¹, twice a day, n = 4 per group) after euthanasia. For PPAR α , FXR, Pol II, p300, CRTC2 (TORC2), NCoR, SMRT, acetyl H4 and H3K27me3 ChIPs (Fig. 4c and Extended Data Figs 9 and 10a), fed or fasted wild-type mice were treated with vehicle, GW7647 (5 mg kg⁻¹) or GW4064 (100 mg kg⁻¹) twice a day (n = 3 per group). Liver tissues (100 mg) were quickly minced using mortar and pestle with liquid nitrogen and cross-linked in 10 ml 1% formaldehyde/PBS for 15 min at room temperature, followed by quenching with 1/20 volume of 2.5 M glycine solution for 5 min. Minced liver tissues were washed twice with cold PBS (150 g, 5 min, 4 °C). Nuclear extracts were prepared by Dounce homogenization (30 strokes on ice, tight fitting pestle-type B) in cell lysis buffer (5 mM PIPES, 85 mM KCl, 0.5% Igepal and complete protease inhibitor tablet from Roche, pH 8.0). Chromatin fragmentation was performed by sonication in 300 μ l ChIP SDS lysis buffer (50 mM HEPES, 1% SDS, 10 mM EDTA, pH 7.5), using the Bioruptor (Diagenode, 4 \times 5 min, 30 s on/30 s off). Proteins of interest were immunoprecipitated in ChIP dilution buffer (50 mM HEPES, 155 mM NaCl, 1.1% Triton X-100, 0.11% sodium deoxycholate, 1 mM phenylmethylsulphonyl fluoride (PMSF) and complete protease inhibitors tablet, pH 7.5) using 5 μ l of each antibody. Cross-linking was reversed overnight at 65 °C, and DNA was isolated using phenol/chloroform/isoamyl alcohol. Precipitated DNA was analysed by qPCR. Primer sequences used for ChIP-qPCR analysis are listed in Supplementary Table 2.

PPAR α ChIP-seq. For ChIP, frozen liver tissues were ground in liquid nitrogen and cross-linked in 1% formaldehyde for 20 min, followed by quenching with 1/20 volume of 2.5 M glycine solution, and two washes with PBS. Cell lysis and chromatin fragmentation were performed by sonication in ChIP dilution buffer (50 mM HEPES, 155 mM NaCl, 1.1% Triton X-100, 1% SDS, 0.11% sodium deoxycholate, 1 mM EDTA, and complete protease inhibitors tablet, pH 7.5). PPAR α proteins were immunoprecipitated using PPAR α antibody, cross-linking was reversed overnight at 65 °C, and DNA was isolated using phenol/chloroform/isoamyl alcohol. Precipitated DNA was analysed by qPCR. ChIP was performed independently on liver samples from fed wild-type or *Ppara*^{-/-} mice treated with vehicle or GW7647 (5 mg kg⁻¹, twice a day), and the precipitated DNA or input DNA samples were pooled. Ten nanograms of the pooled DNA was then amplified according to ChIP Sequencing Sample Preparation Guide provided by Illumina, using adaptor oligonucleotides and primers from Illumina, enzymes from New England Biolabs and the PCR Purification Kit and MinElute Kit from Qiagen. Deep sequencing was performed by the Functional Genomics Core (J. Schug and K. Kaestner) of the Penn Diabetes Endocrinology Research Center using the Illumina Genome Analyzer IIx, and sequences were obtained using the Solexa Analysis Pipeline.

ChIP-seq data analysis. Solexa sequencing reads were mapped to reference genome mm9 using Bowtie³⁵. Peak calling was performed with MACS14 using default settings³⁶. Peak heights were normalized to the total number of uniquely mapped reads and displayed in UCSC genome browser³⁷ as the number of tags per 10 million tags. *De novo* motif analysis was performed with MEME³¹ using 200-bp regions centred at the summit of each peak. To analyse the enrichment of autophagy-related genes near PPAR α -binding sites, all genes containing a PPAR α -binding site (with or without the PPAR α motif DR1) within 20 kb from the transcription start site were compared with a list of autophagy-related genes^{32,38–44} (Supplementary Table 1). *P* values were calculated with hypergeometric test.

Transmission electron microscopy. Anaesthetized mice were perfused with PBS for 3 min, followed immediately by perfusion with 2% PFA plus 2.5% glutaraldehyde in 0.1 M Millonig's phosphate buffer, pH 7.4. Liver tissue was quickly removed and minced in a large drop of cold fixative, then transferred to vials of cold fix and fixed overnight in 4 °C. After washing three times in 0.1 M Millonig's phosphate buffer, the tissue was post-fixed at room temperature in 0.1 M Millonig's phosphate buffer containing 1% osmium tetroxide for 1 h. Post-fixation was followed by three rinses, 5 min each, of 0.1 M Millonig's phosphate buffer, after which all tissues were dehydrated through a gradient series of ethanol, beginning with two 10 min changes of 50% ethanol and ending with three 20 min changes of 100% ethanol from a freshly opened bottle. Liver tissue was en bloc stained with saturated uranyl acetate in 50% ethanol for 1 h during the 50% ethanol dehydration stage. After complete dehydration, the tissue was infiltrated over a period of 2 days with progressively concentrated mixtures of plastic resin and 100% ethanol. Infiltration continued until the tissue reached pure resin. The tissue was then given three changes of pure, freshly made Spurr's plastic resin⁴⁵ for 3 h each, after which the liver tissue was embedded in 00 BEEM capsules⁴⁶ and placed in a 68 °C oven overnight. Thin sections of approximately 70 nm were obtained using an RMC MT6000-XL ultramicrotome (RMC) and a Diatome Ultra45 diamond knife (DiATOME), and collected on 150 hex-mesh copper grids. The sections were counterstained with Reynold's lead citrate⁴⁷ for 4 min. Dry samples were examined on a Hitachi H7500 transmission electron microscope and images were captured using a Gatan US1000 digital camera with Digital Micrograph (Gatan, v1.82.366) software.

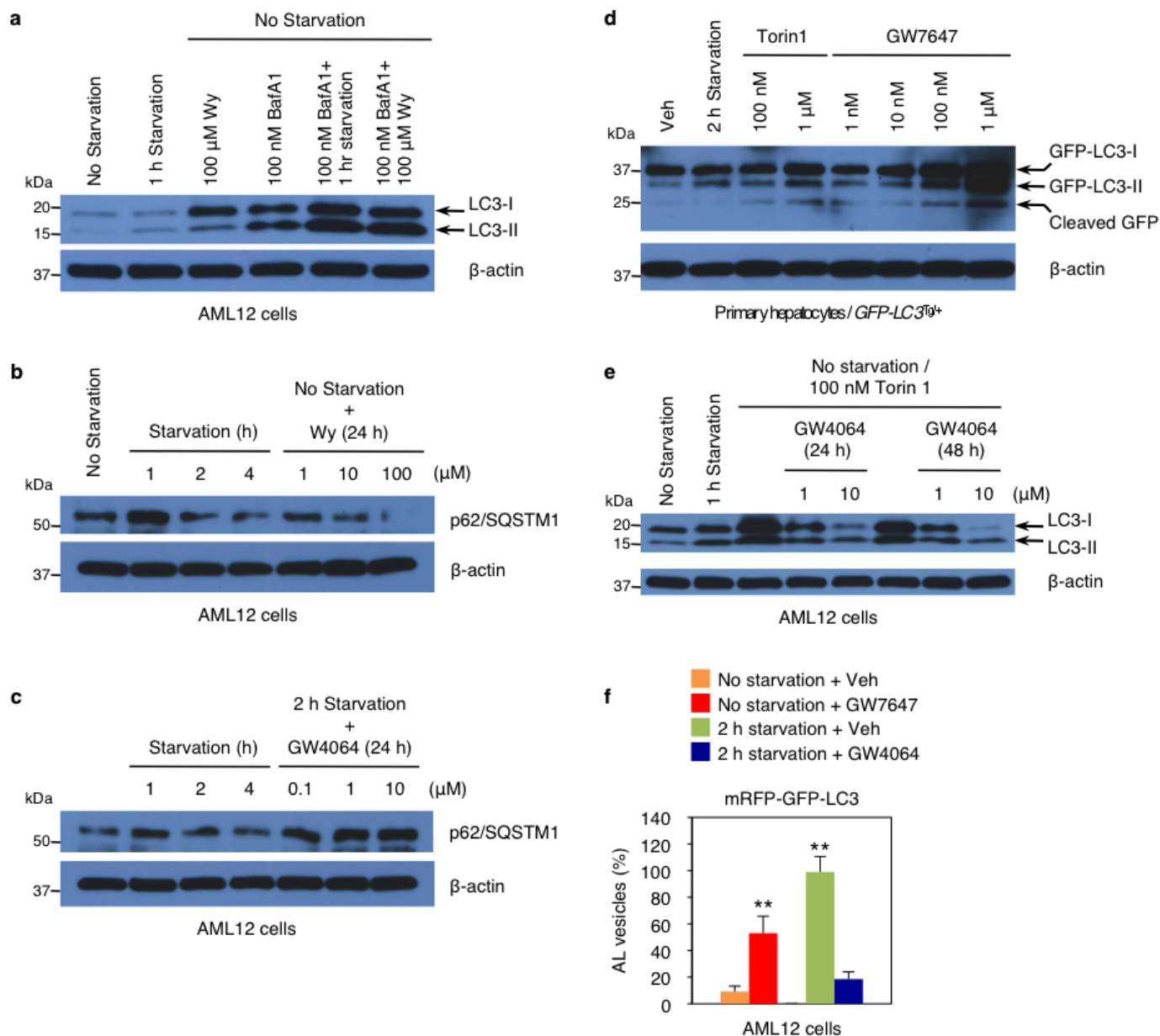
Molecular cloning and cell-based luciferase reporter assay. The oligonucleotides encompassing three copies of DR1 motifs found in mouse LC3a or LC3b enhancer region were annealed and treated with Klenow. Purified DNAs were cloned into pTK-luc plasmid by a serial digesting with SacI and BglII. Successful cloning was confirmed by DNA sequencing analysis. Oligonucleotide sequences used for luciferase reporter assays are listed in Supplementary Table 3. For the luciferase assay, AML12 cells were cultured in 24-well plates. Transient transfections were performed using Lipofectamine 2000. Cells were transfected with 200 ng of reporter constructs, 100 ng of cytomegalovirus-promoter (CMX)-mouse PPAR α , or CMX-human FXR, and 50 ng of CMX- β -galactosidase. pCDNA3.1 was added to prepare the total DNA to 500 ng per well. After 4 h transfection, cells were treated with or vehicle (0.1%

DMSO), 10 μ M Wy-14,643 or 1 μ M GW4064 in media containing charcoal-stripped serum for 20 h before performing luciferase and β -galactosidase assay. Luciferase activity was normalized with β -galactosidase activity. Normalized values from vehicle-treated cells transfected with empty CMX plasmid were set as fold 1.

Measurement of β -hydroxybutyrate. AML12 cells were transfected with siControl (ON-TARGETplus Control pool, D-001810-10-20), siAtg5 (ON-TARGETplus SMARTpool siRNA Atg5, L-064838-00-0010), or siAtg7 (ON-TARGETplus SMARTpool siRNA Atg7, L-049953-00-0010) according to manufacturer's instructions. siRNAs were purchased from Thermo Scientific Dharmacon. Twenty-four hours after transfection using Lipofectamine 2000, cells were treated with vehicle (0.1% DMSO) or Wy-14,643 (10 μ M) in media containing charcoal-stripped serum and 250 μ M oleate for 48 h. Released β -hydroxybutyrate in the medium was measured using a β -hydroxybutyrate assay kit. Female *Atg7*^{F/F} mice²⁴ were bred with male *Alb-Cre/+* mice to generate control littermates (*Atg7*^{F/F}) and hepatocyte-specific *Atg7*-null (*Alb-Cre/+ Atg7*^{F/F}) mice. Eight-week-old male *Atg7*^{F/F} and *Alb-Cre/+ Atg7*^{F/F} mice were orally gavaged with either vehicle (DMSO in 4:1 of PEG-400 and Tween 80) or GW7647 (5 mg kg⁻¹ body weight) twice a day (first injection: 00:00, second injection: 12:00). Mice were fed *ad libitum* or fasted for 24 h from 18:00 (day 1) to 18:00 (day 2) and then sacrificed to collect blood and livers. Serum β -hydroxybutyrate concentrations were normalized with mouse liver weights.

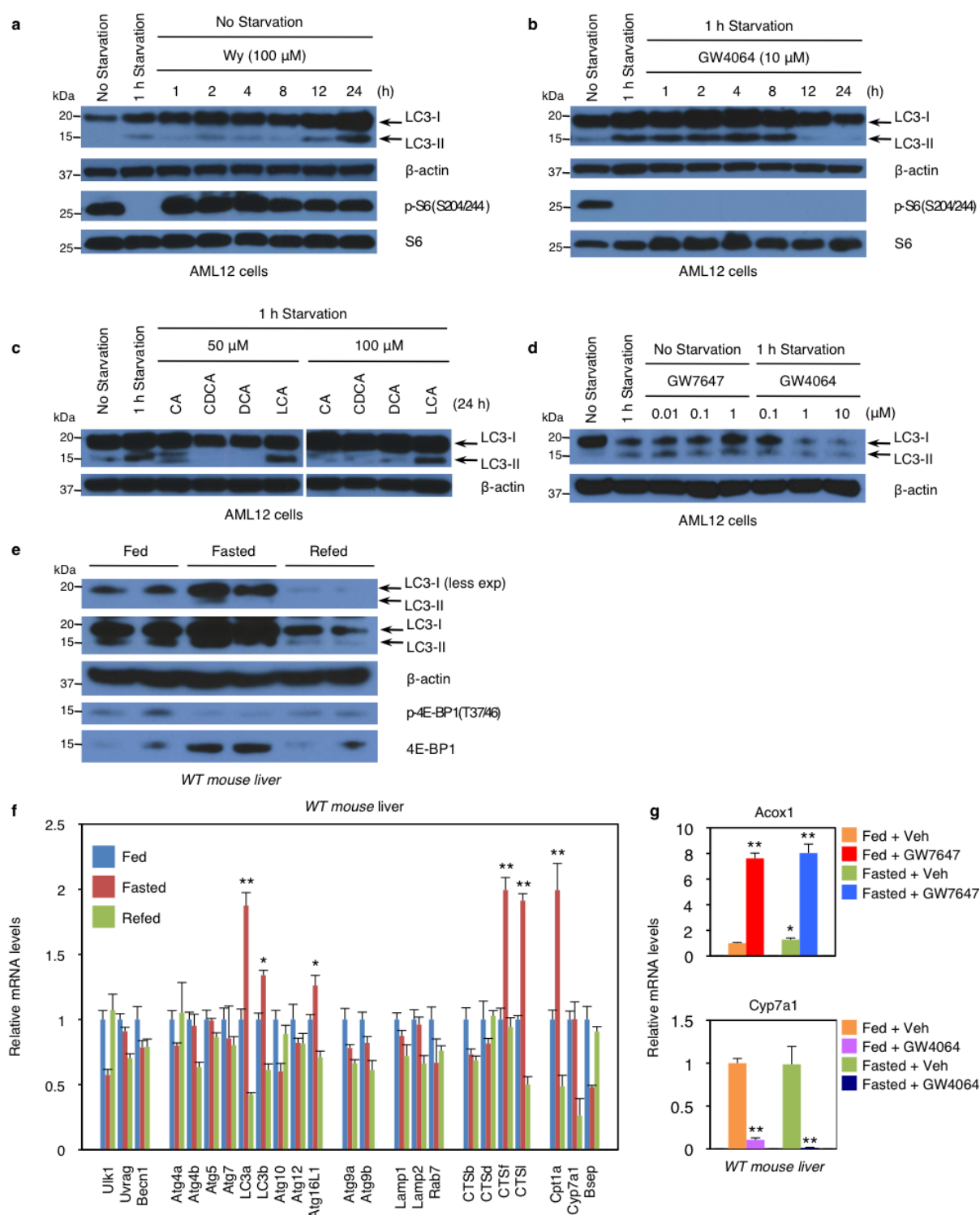
Statistical analysis. Sample size for experiments was determined empirically based on preliminary experiments to ensure appropriate statistical power. Age, sex and weight-matched mice were randomly assigned for the treatments. No animals were excluded from statistical analysis, and the investigators were not blinded in the studies. Values are expressed as mean \pm s.e.m., and error bars for all results were derived from biological replicates rather than technical replicates. Significant differences between two groups were evaluated using a two-tailed, unpaired *t*-test, which was found to be appropriate as groups displayed a normal distribution and comparable variance. *P* < 0.05 was considered statistically significant.

31. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
32. Mizushima, N., Yoshimori, T. & Ohsumi, Y. The role of Atg proteins in autophagosome formation. *Annu. Rev. Cell Dev. Biol.* **27**, 107–132 (2011).
33. Sun, Z. *et al.* Deacetylase-independent function of HDAC3 in transcription and metabolism requires nuclear receptor corepressor. *Mol. Cell* **52**, 769–782 (2013).
34. Feng, D. *et al.* A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. *Science* **331**, 1315–1319 (2011).
35. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
36. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
37. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
38. Kundu, M. & Thompson, C. B. Autophagy: basic principles and relevance to disease. *Annu. Rev. Pathol.* **3**, 427–455 (2008).
39. Jin, M., Liu, X. & Klionsky, D. J. SnapShot: Selective autophagy. *Cell* **152**, 368–368 (2013).
40. Settembre, C. *et al.* TFEB links autophagy to lysosomal biogenesis. *Science* **332**, 1429–1433 (2011).
41. Warr, M. R. *et al.* FOXO3A directs a protective autophagy program in haematopoietic stem cells. *Nature* **494**, 323–327 (2013).
42. Shoji-Kawata, S. *et al.* Identification of a candidate therapeutic autophagy-inducing peptide. *Nature* **494**, 201–206 (2013).
43. Itakura, E., Kishi-Itakura, C. & Mizushima, N. The hairpin-type tail-anchored SNARE syntaxin 17 targets to autophagosomes for fusion with endosomes/lysosomes. *Cell* **151**, 1256–1269 (2012).
44. McKnight, N. C. *et al.* Genome-wide siRNA screen reveals amino acid starvation-induced autophagy requires SCOC and WAC. *EMBO J.* **31**, 1931–1946 (2012).
45. Spurr, A. R. A low-viscosity epoxy resin embedding medium for electron microscopy. *J. Ultrastruct. Res.* **26**, 31–42 (1969).
46. Glauber, A. M. *Practical Methods in Electron Microscopy* 143–144 (North-Holland American Elsevier, 1975).
47. Reynolds, E. S. The use of lead citrate at high pH as an electron-opaque stain in electron microscopy. *J. Cell Biol.* **17**, 208–212 (1963).



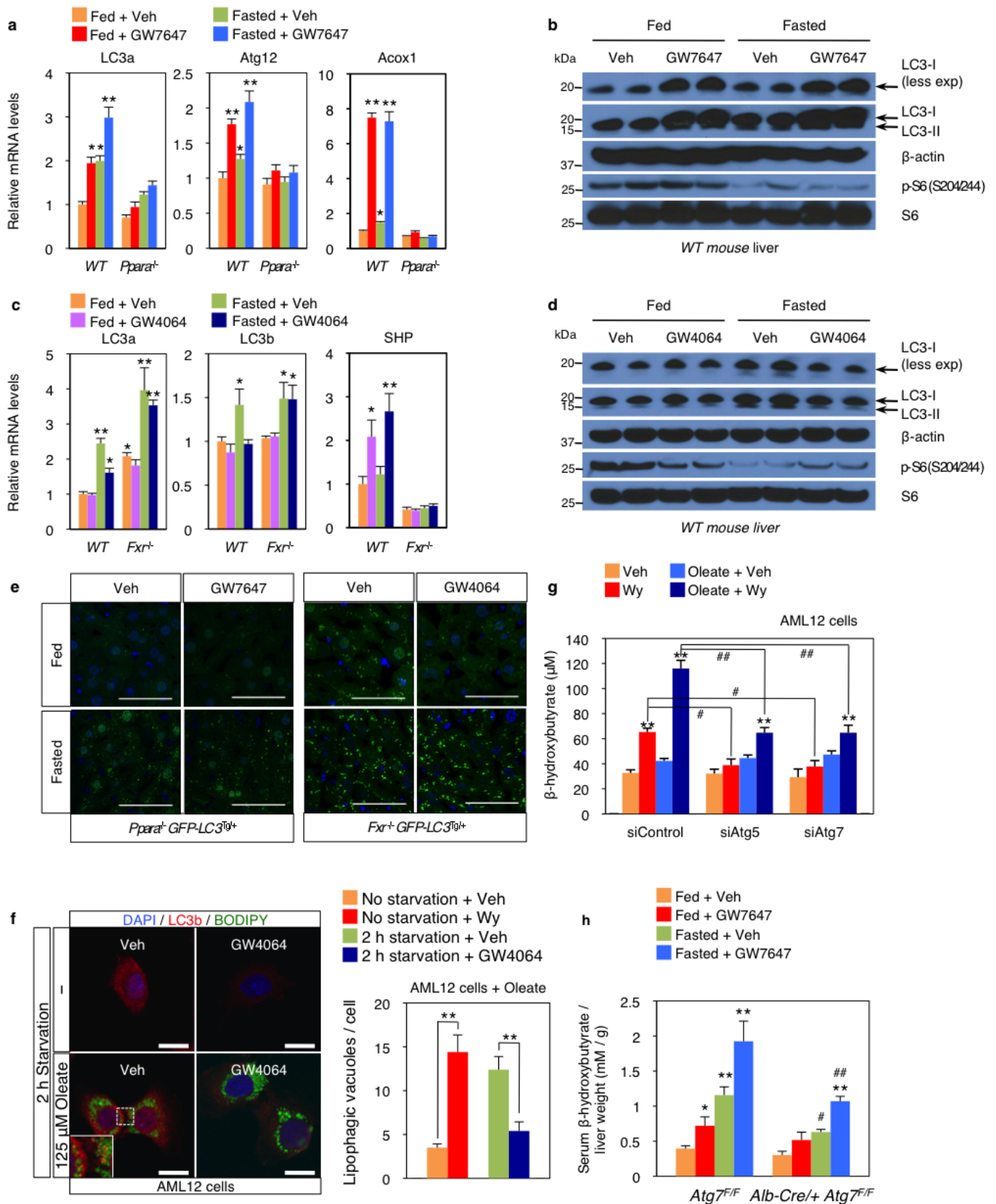
Extended Data Figure 1 | PPAR α or FXR agonist affects autophagic flux in mouse hepatocytes. **a**, Autophagic flux was assessed by LC3 immunoblot analysis in AML12 cells treated with indicated dose of Wy-14,643, or co-treated with both Wy-14,643 and bafilomycin A₁ (BafA1). **b**, **c**, Biochemical determination of autophagy (p62/SQSTM1 immunoblot) in AML12 cells treated with indicated doses of Wy-14,643 or GW4064 for 24 h. GW4064-treated cells were starved for 2 h. **d**, Primary hepatocytes from GFP-LC3^{Tg/+} mice were treated with indicated doses of Torin1 or GW7647. GFP-LC3 cleavage was assessed by GFP immunoblot analysis. **e**, LC3 immunoblot in

AML12 cells treated with indicated doses of GW4064 or co-treated with GW4064 and Torin1. All drug treatments were done in complete media for 24 h unless otherwise indicated. β -actin is a loading control. **f**, Quantification of autophagic flux shown in Fig. 1a. Numbers of autolysosomes (ALs) induced by 2 h starvation plus vehicle were set as 100%. Numbers of autolysosomes = RFP-positive vesicles - GFP-positive vesicles. Thirty cells were counted per condition (** $P < 0.01$ versus no starvation plus vehicle; two-tailed t -test). Data represent mean \pm s.e.m.



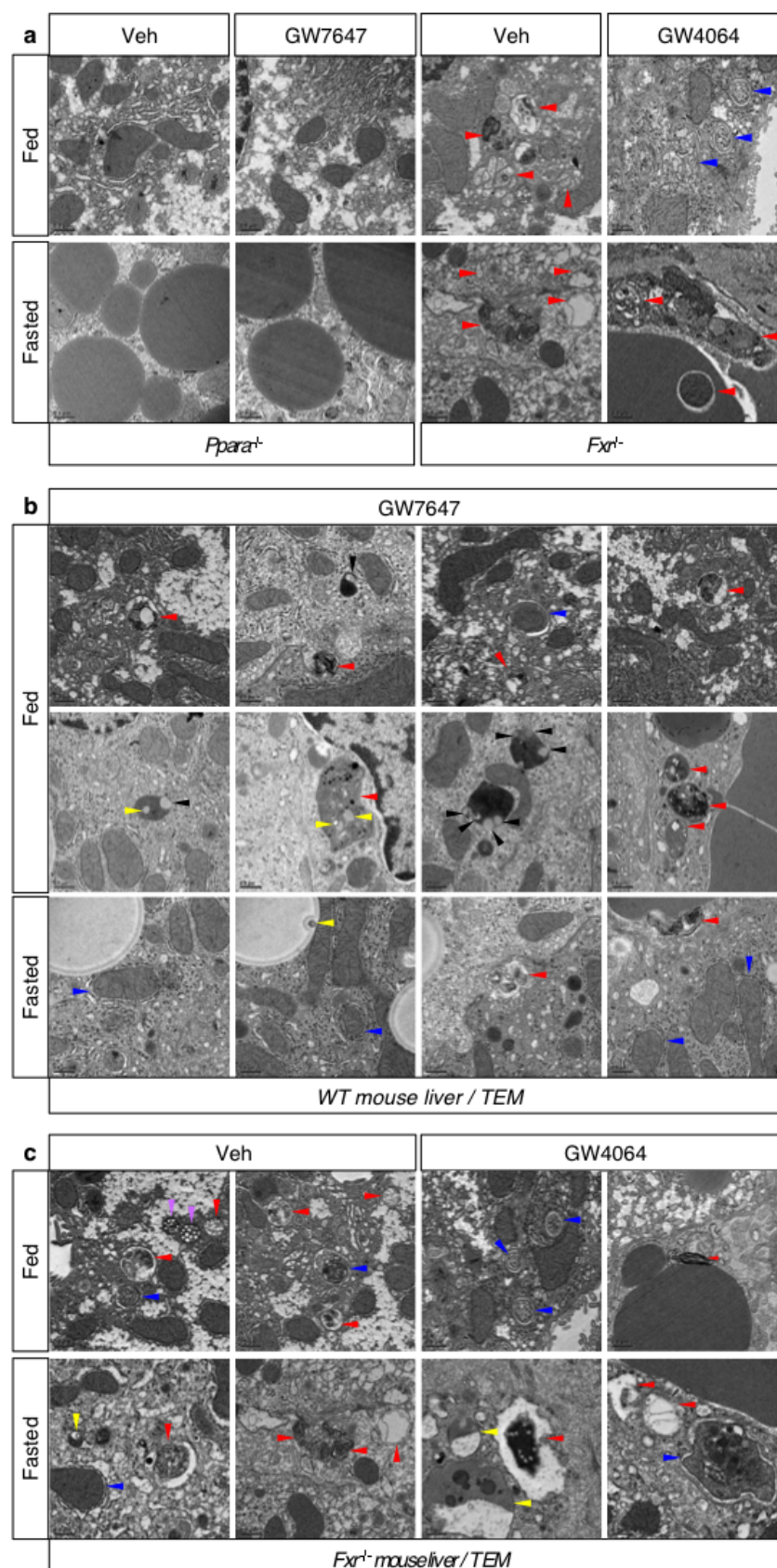
Extended Data Figure 2 | Nutrient availability regulates the expression of autophagy-related genes. **a, b,** Biochemical determination of autophagy (LC3 immunoblot; LC3-I and LC3-II denote non-lipidated and phosphatidylethanolamine-conjugated forms of MAP1LC3, respectively) or mTORC1 activity (p-S6(Ser 240/244) immunoblot) in AML12 cells treated for indicated time with 100 μ M Wy-14,643 or 10 μ M GW4064. GW4064-treated cells were starved in HBSS medium for 1 h. β -actin is a loading control. **c,** Bile acids suppress autophagosome formation. AML12 cells were treated with indicated doses of each bile acid for 24 h, followed by 1-h starvation in HBSS medium. **d,** Immunoblot analysis of LC3-I/II and β -actin in AML12 cells treated with indicated dose of GW7647 or GW4064 for 24 h. GW4064-treated cells were starved in HBSS medium for 1 h. **e, f,** Fed *ad libitum*, 24 h fasted, or 24 h refed after 24 h fasting wild-type mice were euthanized to collect livers at

18:00. **e,** Immunoblot analysis of LC3-I/II, β -actin, phosphorylated 4E-BP1 (p-4E-BP1(Thr 37/46)) and total 4E-BP1 in liver samples. Proteins (50 μ g) from liver homogenates were separated by SDS-PAGE and probed with the indicated antibodies. β -actin is a loading control. A pooled sample was loaded onto the gel in duplicate ($n = 5$ per group). **f,** Hepatic expression levels of autophagy-related genes, and PPAR α and FXR target genes affected by nutrient availability ($n = 5$ per group, * $P < 0.05$, ** $P < 0.01$ versus fed wild-type mice; two-tailed *t*-test). **g,** Hepatic expression levels of PPAR α or FXR target gene (*Acox1* and *Cyp7a1*, respectively) were determined by qPCR analysis. Fed or fasted wild-type mice were orally gavaged with vehicle, GW7647 or GW4064 twice a day ($n = 5$ per group, * $P < 0.05$, ** $P < 0.01$ versus fed wild-type mice). Data represent mean \pm s.e.m.



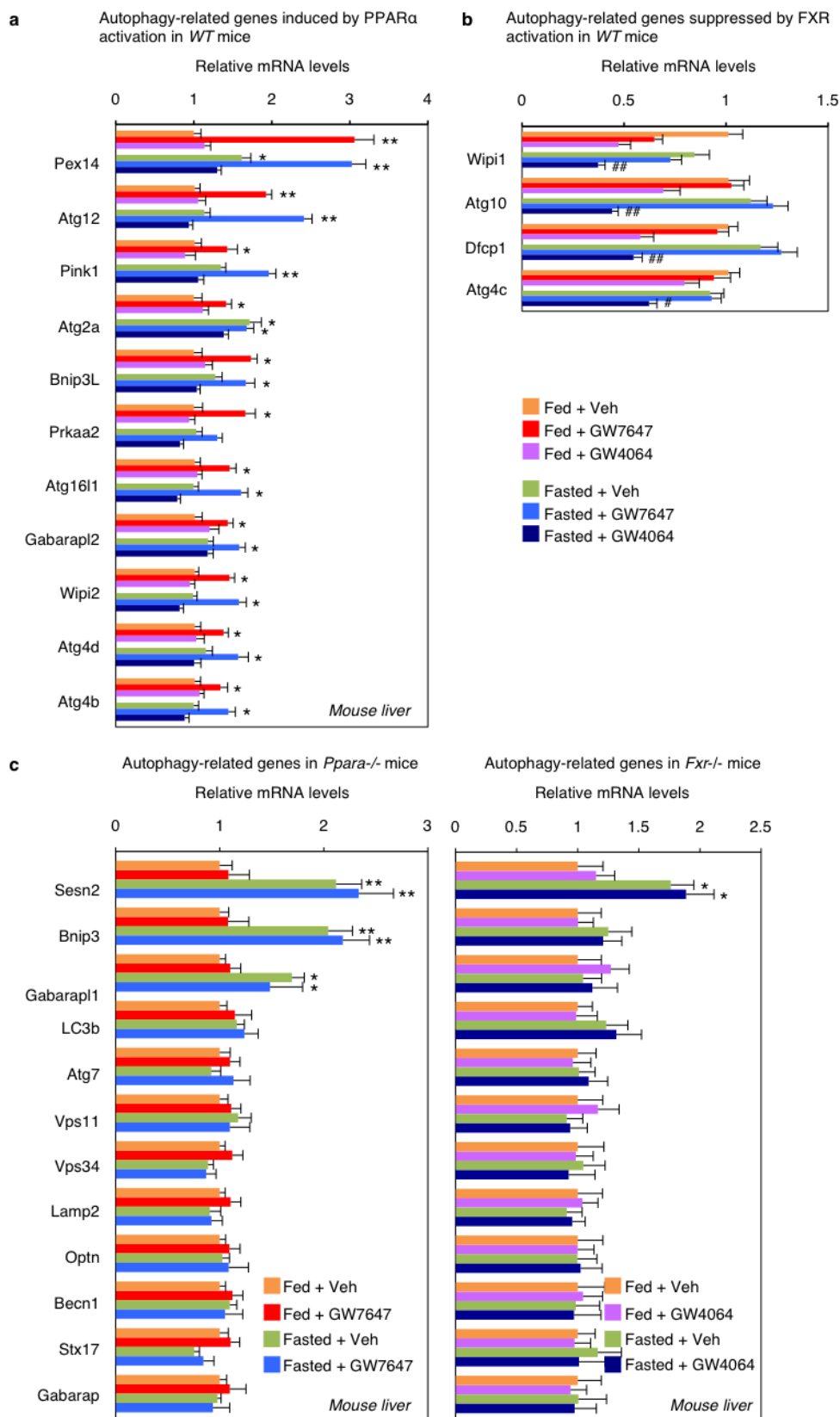
Extended Data Figure 3 | Pharmacological activation of PPAR α or FXR in fed or fasted mouse liver. **a, c,** Hepatic expression levels of autophagy-related genes (*LC3a*, *LC3b* and *Atg12*), PPAR α -target gene (*Acox1*) and FXR-target gene (*SHP*) were determined by qPCR analysis. Fed or fasted wild-type, *Ppara*^{-/-} or *Fxr*^{-/-} mice were orally treated with vehicle, GW7647 or GW4064 twice a day ($n = 5$ per group, * $P < 0.05$, ** $P < 0.01$ versus fed wild-type mice treated with vehicle). **b, d,** Immunoblot analysis of LC3-I/II, β -actin, phosphorylated S6 (p-S6) and total S6 in liver samples. Fed or fasted wild-type mice were orally gavaged with vehicle, GW7647 or GW4064 twice a day. A pooled sample was loaded onto the gel in duplicates ($n = 5$ per group). β -actin is a loading control. **e,** Representative confocal images (out of nine tissue sections per condition) of GFP-LC3 puncta (green: autophagosomes) and DAPI (blue: DNA) staining in livers. Fed or fasted bigenic *Ppara*^{-/-} GFP-LC3^{Tg/+} or *Fxr*^{-/-} GFP-LC3^{Tg/+} mice were orally gavaged with vehicle, GW7647 or GW4064 twice a day. Liver samples were fixed and cryosections were analysed by confocal microscopy. Scale bars, 50 μ m. **f,** Co-localization of BODIPY 493/503 (green) with LC3 (red) in AML12 cells treated with vehicle or 1 μ M GW4064 for 24 h and simultaneously cultured with or without 125 μ M

oleate in complete medium. GW4064-treated cells were starved for 2 h. DNA was stained with DAPI (blue). Scale bars, 20 μ m. Quantification of lipophagic vacuoles shown in **f** and Fig. 3b. Thirty cells were counted per condition (** $P < 0.01$). **g,** Measuring β -hydroxybutyrate. AML12 cells were transiently transfected with control siRNA (siControl), *Atg5* siRNA (siAtg5) or *Atg7* siRNA (siAtg7) for 24 h followed by indicated drug treatments for 48 h with or without 250 μ M oleate (vehicle: 0.1% DMSO, Wy: 10 μ M Wy-14,643). Released β -hydroxybutyrate in the medium was determined (** $P < 0.01$ versus siControl treated with vehicle; # $P < 0.01$ versus siControl treated with Wy-14,643; ## $P < 0.01$ versus siControl treated with oleate plus Wy-14,643). **h,** Serum β -hydroxybutyrate were normalized to liver weights. Fed or 24-h fasted control littermates (*Atg7*^{F/F}) and hepatocyte-specific *Atg7*^{F/F}-null (*Alb-Cre/+ Atg7*^{F/F}) mice were treated with vehicle or GW7647 twice a day ($n = 4$ per group, * $P < 0.05$, ** $P < 0.01$ versus fed *Atg7*^{F/F} mice treated with vehicle; # $P < 0.01$ versus fasted *Atg7*^{F/F} mice treated with vehicle; ## $P < 0.01$ versus fasted *Atg7*^{F/F} mice treated with GW7647). Data are mean \pm s.e.m. Statistics by two-tailed *t*-test.



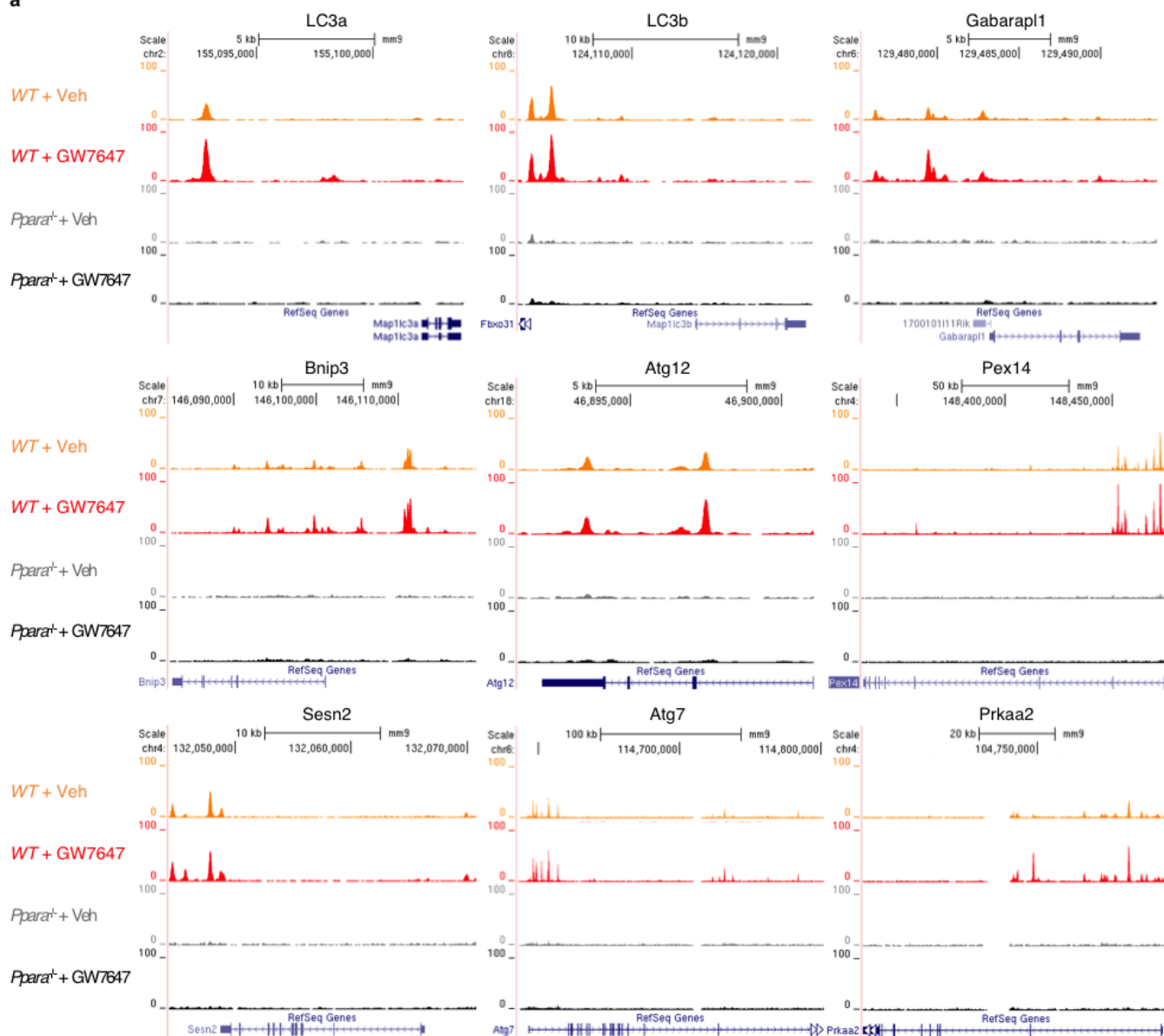
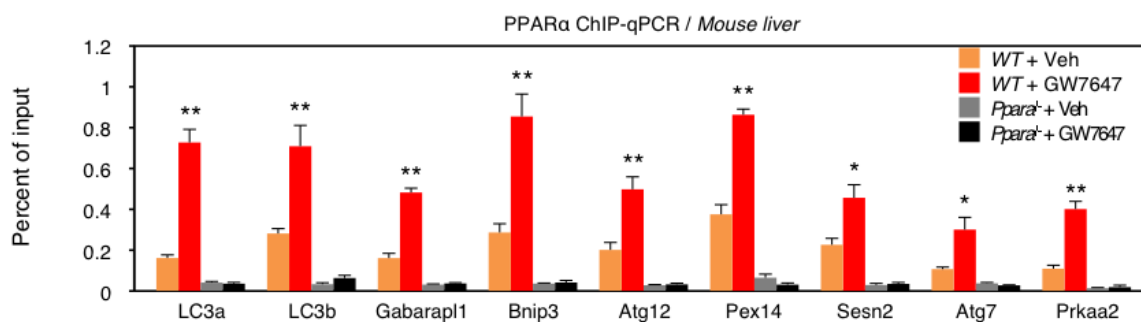
Extended Data Figure 4 | PPAR α activation or loss of FXR induces autophagy in liver. Magnification of representative transmission electron micrograph images (out of 30 cells per group) of livers. **a–c**, Fed or fasted wild-type, *Ppara*^{-/-} or *Fxr*^{-/-} mice were orally gavaged with vehicle, GW7647

or GW4064 twice a day. Lipophagy (yellow arrowheads), autophagosomes (blue arrowheads), autolysosomes (red arrowheads), microautophagy (black arrowheads) and multivesicular bodies (purple arrowheads). Scale bars, 0.5 μ m.



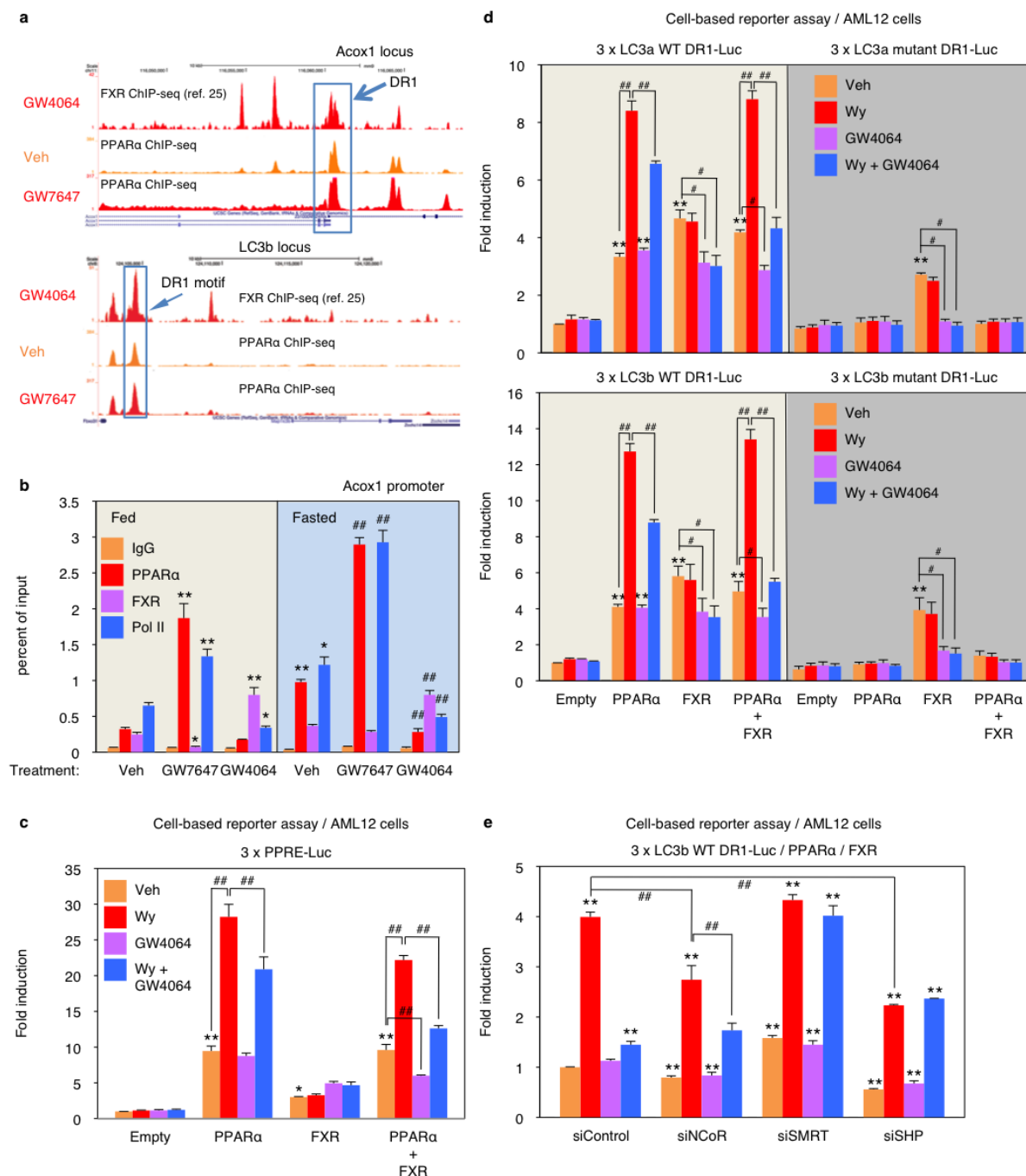
Extended Data Figure 5 | Expression profiles of autophagy-related genes by PPAR α or FXR activation in liver. a–c, Hepatic expression levels of autophagy-related genes were determined by qPCR analysis in wild-type (a, b), *Ppara*^{-/-} or *Fxr*^{-/-} (c) mice. Eleven genes in a were induced by PPAR α activation, but not affected by FXR activation. Four genes in b were suppressed by FXR activation, but not affected by PPAR α activation (a and b, $n = 5$ per group, * $P < 0.05$, ** $P < 0.01$ versus fed wild-type mice treated with vehicle;

$P < 0.05$, ## $P < 0.01$ versus fasted wild-type mice treated with vehicle). Altered expression levels of 13 genes shown in Fig. 4a were lost in *Ppara*^{-/-} or *Fxr*^{-/-} mice in c ($n = 5$ per group, * $P < 0.05$, ** $P < 0.01$ versus fed *Ppara*^{-/-} or *Fxr*^{-/-} mice treated with vehicle). Fed or fasted wild-type, *Ppara*^{-/-} or *Fxr*^{-/-} mice were orally gavaged with vehicle, GW7647 or GW4064 twice a day. Data represent mean \pm s.e.m. Statistics by two-tailed t -test.

a**b**

Extended Data Figure 7 | PPARα ChIP-seq profiles at loci of autophagy-related genes. Fed wild-type or *Ppara*^{-/-} mice were orally gavaged with vehicle or GW7647 twice a day. Mouse livers were taken out 6 h after the last injection of drugs to perform PPARα ChIP-seq and ChIP-qPCR. **a**, Representative ChIP-seq reads for PPARα aligned to the autophagy-related

genes (*LC3a*, *LC3b*, *Gabarapl1*, *Bnip3*, *Atg12*, *Pex14*, *Sesn2*, *Atg7* and *Prkaa2*). **b**, PPARα ChIP-qPCR for autophagy-related genes shown in **a** ($n = 4$ per group, * $P < 0.05$, ** $P < 0.01$ versus fed wild-type mice treated with vehicle; two-tailed t -test). Data represent mean \pm s.e.m.

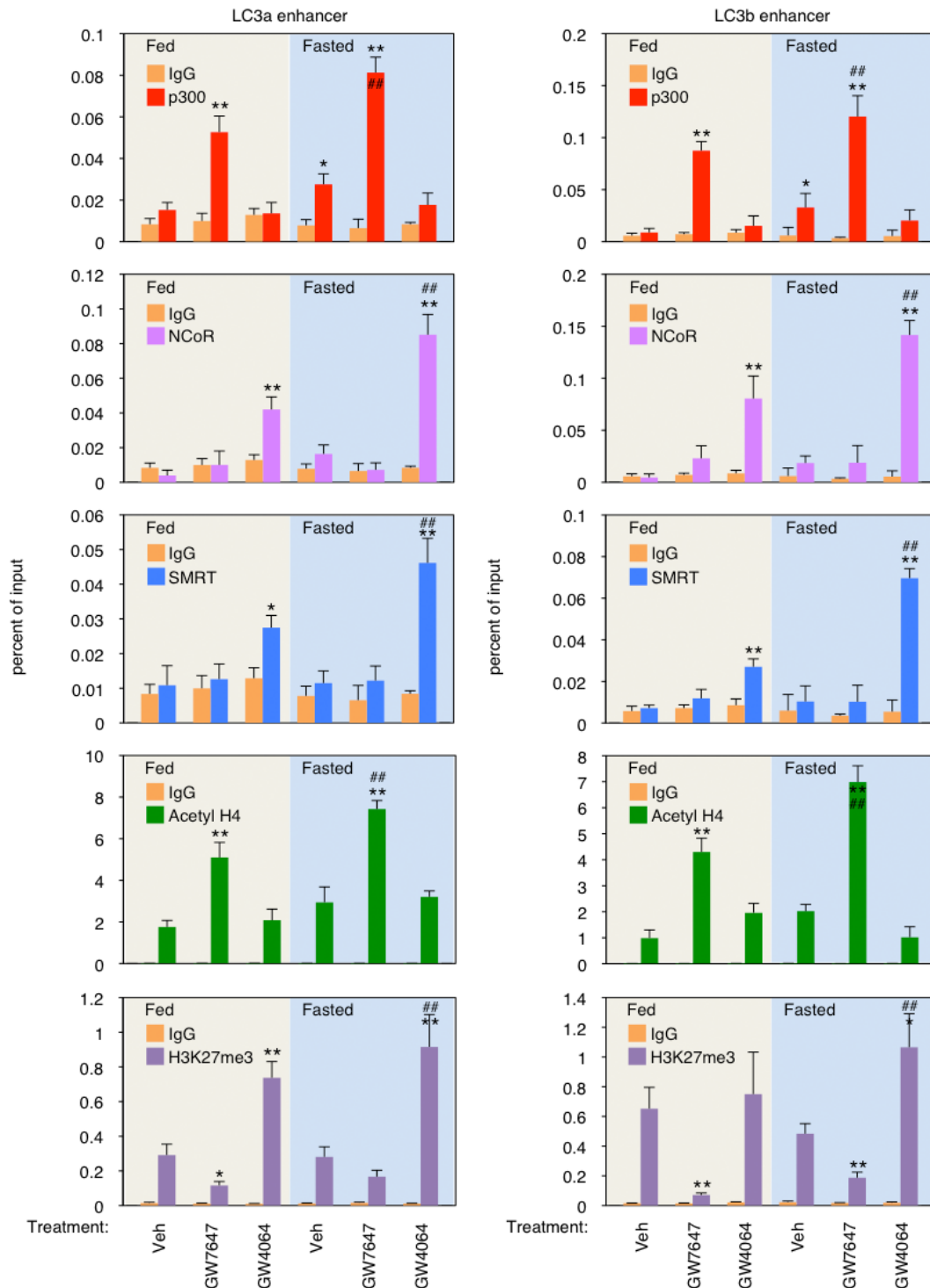


Extended Data Figure 8 | PPAR α /FXR genomic competition for DR1 in *Acox1* gene and autophagy-related genes.

a, Representative ChIP-seq reads for FXR and PPAR α aligned to the *Acox1* and *LC3b* genes. The peaks in the box contain DR1 motif. Fed wild-type mice were orally gavaged with vehicle or GW764 twice a day ($n = 4$ per group). **b**, PPAR α or FXR ChIP-qPCR in livers. Fed or fasted wild-type mice were orally gavaged with vehicle, GW7647 or GW4064 twice a day ($n = 3$ per group, $*P < 0.05$, $**P < 0.01$ versus fed wild-type mice treated with vehicle; $##P < 0.01$ versus fasted wild-type mice treated with vehicle). **c**, Cell-based luciferase reporter assays. AML12 cells were transiently transfected with a 3 \times PPRE luciferase reporter construct (3 \times PPRE-luc) and CMX- β -galactosidase in a combination of expression plasmids of PPAR α , FXR or both, followed by drug treatment for 20 h (vehicle: 0.1% DMSO; Wy: 10 μ M Wy-14,643; GW4064: 1 μ M GW4064). Normalized values (luciferase activity/ β -galactosidase activity) of vehicle-treated cells transfected with empty plasmid were set as fold 1 ($*P < 0.05$, $**P < 0.01$ versus empty treated with vehicle; $##P < 0.01$). **d**, Functional role of DR1 motif in the regulatory region of mouse LC3a and LC3b for PPAR α or FXR activity. Cell-based luciferase reporter assays were performed in AML12 cells by

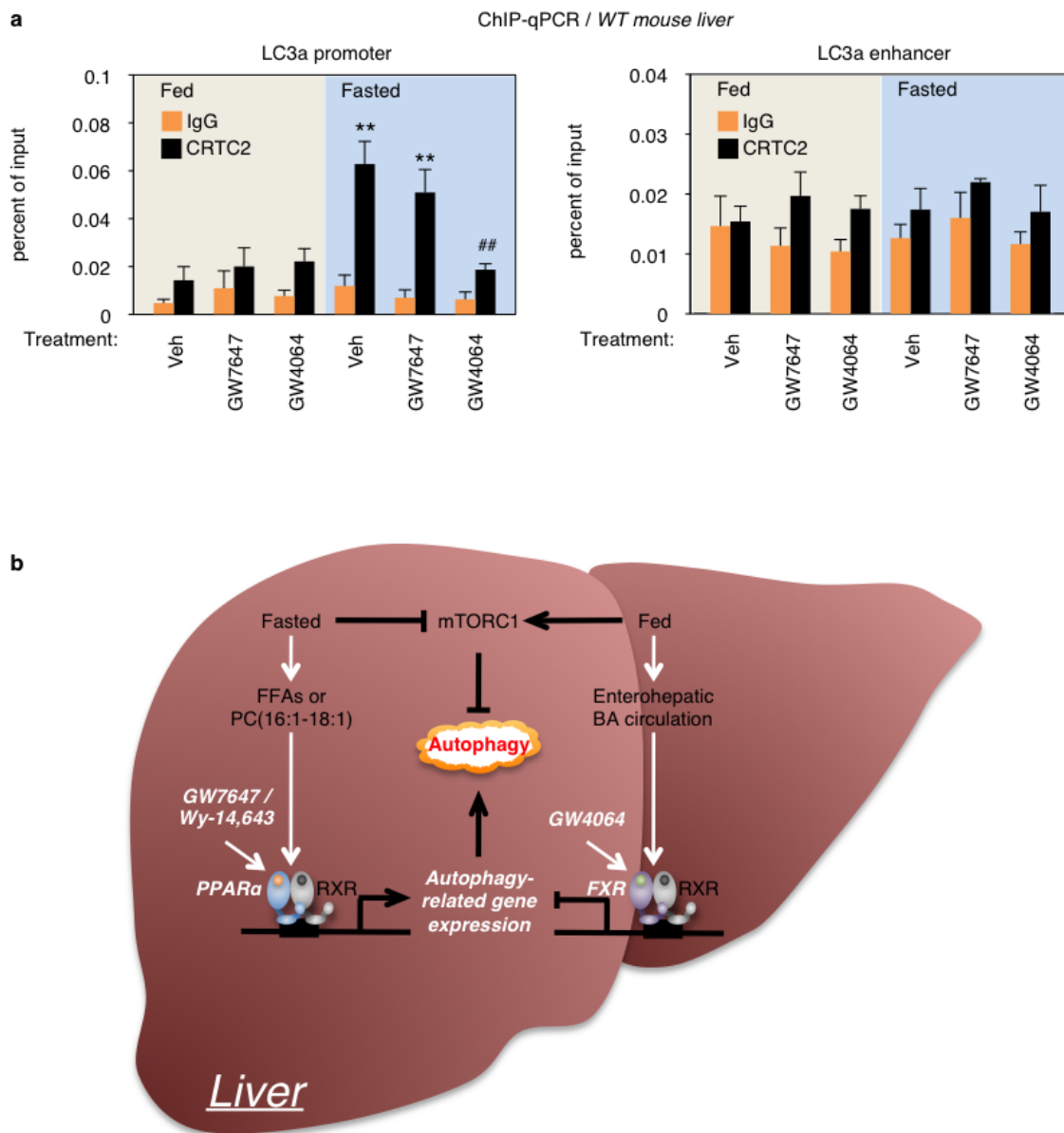
transiently transfecting three tandem copies of mouse LC3a/LC3b DR1 luciferase reporter construct (3 \times LC3a/LC3b DR1 WT-luc) or mutated version (3 \times LC3a/LC3b DR1 mutant-luc) and CMX- β -galactosidase in a combination of expression plasmids of PPAR α , FXR or both, followed by drug treatment for 20 h (vehicle: 0.1% DMSO; Wy: 10 μ M Wy-14,643; GW4064: 1 μ M GW4064). Normalized values (luciferase activity/ β -galactosidase activity) of vehicle-treated cells transfected with empty plasmid were set as fold 1 ($*P < 0.05$, $**P < 0.01$ versus empty treated with vehicle; $##P < 0.01$). **e**, Cell-based luciferase reporter assays were performed in AML12 cells by transiently transfecting siControl, siNCoR, siSMRT or siSHP along with three tandem copies of mouse LC3b DR1 luciferase reporter construct (3 \times LC3b DR1 WT-luc), expression plasmids of PPAR α and FXR, and CMX- β -galactosidase followed by drug treatment for 24 h (vehicle: 0.1% DMSO; Wy: 10 μ M Wy-14,643; GW4064: 1 μ M GW4064). Normalized values (luciferase activity/ β -galactosidase activity) of vehicle-treated cells transfected with siControl were set as fold 1 ($*P < 0.05$, $**P < 0.01$ versus siControl treated with vehicle; $##P < 0.01$). Data represent mean \pm s.e.m. Statistics by two-tailed t -test.

ChIP-qPCR / WT mouse liver



Extended Data Figure 9 | PPAR α or FXR activation controls recruitments of co-regulators and epigenetic marks in the enhancer regions of *LC3a* and *LC3b* genes. Fed or fasted wild-type mice were orally gavaged with vehicle, GW7647 or GW4064 twice a day. Hepatic ChIP-qPCR analysis with indicated antibodies (p300, NCoR1, SMRT, acetyl-H4 and H3K27me3) was used to determine recruitments of co-regulators and subsequent alterations of

epigenetic marks induced by PPAR α /FXR genomic competition for DR1 found in the enhancer region of *LC3a* and *LC3b* genes ($n = 3$ per group, $*P < 0.05$, $**P < 0.01$ versus fed wild-type mice treated with vehicle; $##P < 0.01$ versus fasted wild-type mice treated with vehicle). Data represent mean \pm s.e.m. Statistics by two-tailed t -test.



Extended Data Figure 10 | Working model of the coordination of hepatic autophagy by nutrient-sensing nuclear receptors, PPAR α and FXR.

a, Fed or fasted wild-type mice were orally gavaged with vehicle, GW7647 or GW4064 twice a day. Hepatic CRTC2 ChIP-qPCR in the promoter and enhancer region of *LC3a* gene ($n = 3$ per group, $**P < 0.01$ versus fed wild-type mice treated with vehicle; $##P < 0.01$ versus fasted wild-type mice

treated with vehicle). Data represent mean \pm s.e.m. Statistics by two-tailed *t*-test. **b**, Proposed model depicting transcriptionally activating or suppressive nutrient-sensing nuclear receptors, PPAR α or FXR, respectively, which coordinate autophagy in liver. Activated PPAR α or FXR compete with each other for binding to response elements found in autophagy-related genes.

TRIM37 is a new histone H2A ubiquitin ligase and breast cancer oncoprotein

Sanchita Bhatnagar^{1,2}, Claude Gazin³, Lynn Chamberlain^{1,2}, Jianhong Ou², Xiaochun Zhu^{1,2}, Jogender S. Tushir⁴, Ching-Man Virbasius^{1,2}, Ling Lin^{1,2}, Lihua J. Zhu^{2,5}, Narendra Wajapeyee⁶ & Michael R. Green^{1,2}

The *TRIM37* (also known as *MUL*) gene is located in the 17q23 chromosomal region, which is amplified in up to ~40% of breast cancers¹. *TRIM37* contains a RING finger domain, a hallmark of E3 ubiquitin ligases², but its protein substrate(s) is unknown. Here we report that *TRIM37* mono-ubiquitinates histone H2A, a chromatin modification associated with transcriptional repression³. We find that in human breast cancer cell lines containing amplified 17q23, *TRIM37* is upregulated and, reciprocally, the major H2A ubiquitin ligase RNF2 (also known as RING1B)^{3,4} is downregulated. Genome-wide chromatin immunoprecipitation (ChIP)-chip experiments in 17q23-amplified breast cancer cells identified many genes, including multiple tumour suppressors, whose promoters were bound by *TRIM37* and enriched for ubiquitinated H2A. However, unlike RNF2, which is a subunit of polycomb repressive complex 1 (PRC1)^{3–5}, we find that *TRIM37* associates with polycomb repressive complex 2 (PRC2). *TRIM37*, PRC2 and PRC1 are co-bound to specific target genes, resulting in their transcriptional silencing. RNA-interference-mediated knockdown of *TRIM37* results in loss of ubiquitinated H2A, dissociation of PRC1 and PRC2 from target promoters, and transcriptional reactivation of silenced genes. Knockdown of *TRIM37* in human breast

cancer cells containing amplified 17q23 substantially decreases tumour growth in mouse xenografts. Conversely, ectopic expression of *TRIM37* renders non-transformed cells tumorigenic. Collectively, our results reveal *TRIM37* as an oncogenic H2A ubiquitin ligase that is overexpressed in a subset of breast cancers and promotes transformation by facilitating silencing of tumour suppressors and other genes.

We previously identified *TRIM37* as a factor required for epigenetic silencing of the *Fas* tumour suppressor gene, and found that *TRIM37* was bound to the silenced *Fas* promoter in *Kras*-transformed NIH 3T3 cells⁶. Mono-ubiquitination of histone H2A at lysine 119 is associated with transcriptional repression³ and we considered the possibility that *TRIM37* was a histone H2A E3 ubiquitin ligase.

To test this idea, we first performed ChIP experiments to assess ubiquitinated H2A (H2Aub) levels on the epigenetically repressed *Fas* promoter. H2Aub was enriched at the silenced *Fas* promoter in *Kras*-transformed NIH 3T3 cells, and was lost after short hairpin RNA (shRNA)-mediated knockdown of *TRIM37* (Extended Data Fig. 1a–c). By contrast, H2Aub was retained after knockdown of RNF2, the major H2A ubiquitin ligase involved in gene regulation^{3,4}, or BMI1, a PRC1 component^{7,8} that stimulates RNF2 activity^{7,9}.

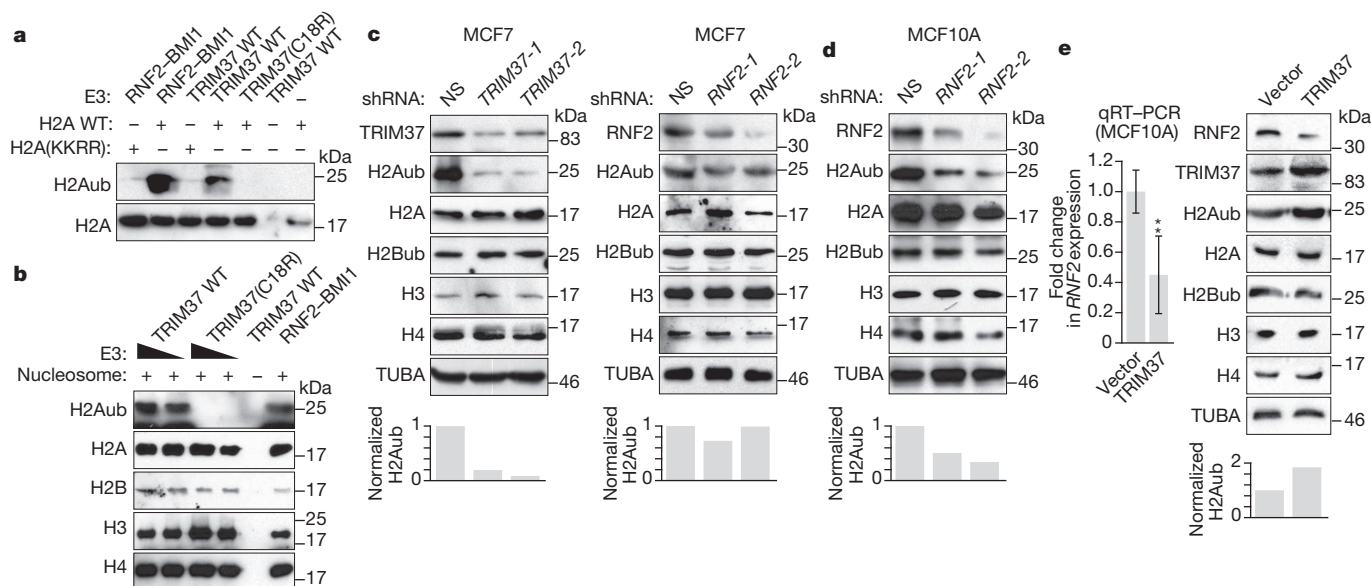


Figure 1 | *TRIM37* is a histone H2A ubiquitin ligase that is overexpressed in 17q23-amplified human breast cancer cell lines. **a, b,** *In vitro* ubiquitination assays with H2A wild type (WT) or H2A(KKRR) (**a**), or reconstituted nucleosomes (**b**). **c,** Top, immunoblots in MCF7 cells expressing a non-silencing (NS), *TRIM37* or *RNF2* shRNA. Bottom, quantification of H2Aub relative to α -tubulin (TUBA); H2Aub in NS cells was set to 1. **d,** Immunoblots

in MCF10A cells expressing an NS or *RNF2* shRNA. **e,** qRT-PCR (left) and immunoblots (right) in MCF10A cells expressing *TRIM37* or vector. *RNF2* expression and H2Aub levels were normalized to that obtained with vector. Error bars indicate standard error of the mean (s.e.m.); $n = 3$ technical replicates of a representative experiment (out of three experiments). * $P < 0.05$; ** $P < 0.01$.

¹Howard Hughes Medical Institute, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. ²Programs in Gene Function and Expression and Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. ³CEA/DSV/IRCM/LEFG, Genopole G2, and Université Paris Diderot, 91057 Evry, France. ⁴Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, Connecticut 06877, USA. ⁵Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. ⁶Department of Pathology, Yale University School of Medicine, New Haven, Connecticut 06520, USA.

We next assessed whether TRIM37 could mono-ubiquitinate H2A at lysine 119. Purified recombinant TRIM37 or, as a positive control, the RNF2–BMI1 complex, was incubated in the presence of an E1 ubiquitin-activating enzyme, an E2 ubiquitin-conjugating enzyme, ATP and haemagglutinin (HA)-tagged ubiquitin. Like RNF2–BMI1, TRIM37 could mono-ubiquitinate purified H2A, but not an H2A derivative mutated at the K118/K119 ubiquitination site (H2A(KKRR))¹⁰ or H2B (Fig. 1a and Extended Data Fig. 1d). By contrast, a TRIM37 derivative bearing a point mutation in a conserved cysteine residue in the RING finger motif (C18R), which interferes with catalytic activity², failed to promote H2A mono-ubiquitination (Fig. 1a). Furthermore, wild-type TRIM37, but not TRIM37(C18R), could mono-ubiquitinate H2A in an assembled nucleosome (Fig. 1b).

We next analysed the relationship between expression of TRIM37 and RNF2 in several human breast cancer cell lines containing amplified 17q23 (MCF7, BT474 and MDA-MB-361)¹¹. As controls, we analysed in parallel a non-transformed human breast epithelial cell line that lacked 17q23 amplification (MCF10A) and immortalized human mammary epithelial cells (HMECs). Breast cancer cell lines containing amplified 17q23 expressed substantially higher levels of TRIM37, but lower levels of RNF2, compared with MCF10A cells and HMECs (Extended Data Fig. 1e–g).

Consistent with these expression results, knockdown of TRIM37 in MCF7, BT474 or MDA-MD-361 cells markedly decreased total H2Aub levels (Fig. 1c, left, and Extended Data Fig. 1h–j). By contrast, knockdown of RNF2 (Extended Data Fig. 1k) in MCF7 cells had only a modest effect on H2Aub levels (Fig. 1c, right), but substantially reduced H2Aub levels in MCF10A cells (Fig. 1d). Moreover, knockdown of TRIM37 did not have a major effect on H2Aub levels in breast cancer cell lines lacking 17q23 amplification (Extended Data Fig. 1l, m).

The expression results shown in Extended Data Fig. 1e, f suggested a reciprocal relationship between TRIM37 and RNF2 expression levels. Consistent with this possibility, ectopic expression of TRIM37 in MCF10A cells (Extended Data Fig. 1n) resulted in downregulation of RNF2 (Fig. 1e).

To identify TRIM37 target genes in MCF7 cells, we performed global ChIP–chip analysis and identified 9,412 genes that were bound by TRIM37, 7,638 of which were also enriched for H2Aub (Fig. 2a, Extended Data Fig. 2a and Supplementary Table 1). Knockdown of TRIM37 resulted in loss or reduction of H2Aub in ~75% of these genes (Fig. 2b, Extended Data Fig. 2a, b and Supplementary Table 1). Searches of cancer-gene databases revealed that 469 TRIM37 target genes are putative tumour suppressors (Supplementary Table 1). Notably, there was a statistically significant enrichment of putative tumour suppressors among the genes bound by TRIM37 ($P = 1.99 \times 10^{-5}$).

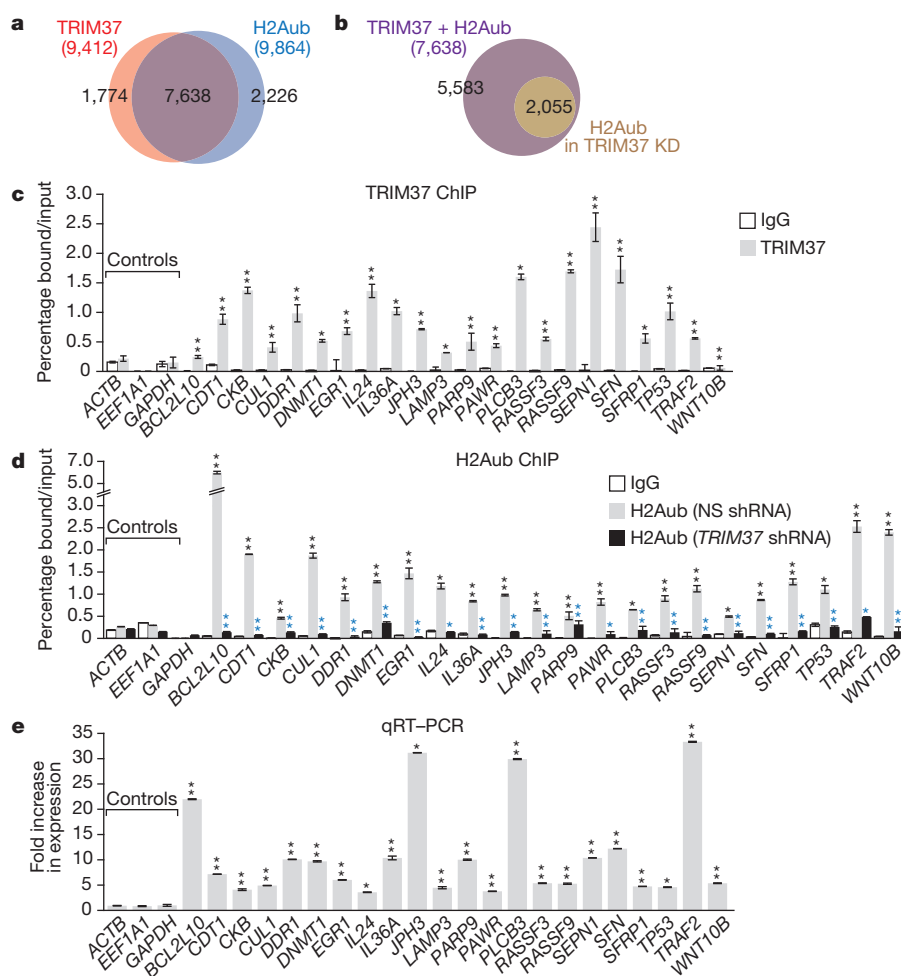


Figure 2 | Identification of TRIM37 target genes. **a, b**, Venn diagrams showing overlap between TRIM37-bound and H2Aub-enriched genes in parental MCF7 cells (**a**) or after TRIM37 knockdown (KD; **b**). See also Supplementary Table 1. **c**, TRIM37 ChIP in MCF7 cells. *ACTB*, *EEF1A1* and *GAPDH* are negative controls. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **d**, H2Aub ChIP in MCF7 cells expressing an NS or TRIM37 shRNA. IgG from cells expressing an NS shRNA is shown. Error bars indicate standard deviation

(s.d.); $n = 3$ technical replicates of a representative experiment (out of three experiments). Black asterisks indicate significance of H2Aub enrichment compared with IgG; blue asterisks indicate significant differences in H2Aub enrichment in cells expressing a TRIM37 relative to an NS shRNA. **e**, qRT–PCR monitoring target gene expression in TRIM37 knockdown MCF7 cells. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). * $P < 0.05$; ** $P < 0.01$.

To validate the ChIP-chip results, we analysed a representative set of 22 putative tumour suppressor and other genes in directed ChIP experiments using promoter-specific primer pairs. For all 22 genes analysed, TRIM37 and H2Aub were significantly enriched at the promoter (Fig. 2c, d and Extended Data Fig. 3a, b). For the majority of these genes, enrichment of H2Aub substantially decreased after knockdown of TRIM37 but not RNF2 (Fig. 2d and Extended Data Fig. 3b, c). Finally, knockdown of TRIM37 but not RNF2 substantially increased expression of all TRIM37 target genes analysed (Fig. 2e and Extended Data Fig. 3d, e). Knockdown of TRIM37 in BT474 cells also resulted in increased expression of a subset of TRIM37 target genes analysed (Extended Data Fig. 3f–h).

Previous studies have shown that RNF2 functions through association with PRC1 (refs 3–5). We next performed a series of experiments to determine whether TRIM37 also functioned through PRCs. We first asked whether TRIM37 was physically associated with PRC1 or PRC2. To test this idea, MCF7 cell nuclear extract was fractionated by sucrose gradient sedimentation and individual fractions were analysed by immunoblotting for TRIM37 and PRC subunits. The results shown in Fig. 3a demonstrate that TRIM37 co-sedimented with PRC2 subunits EZH2 and SUZ12 (refs 8, 12, 13). Physical interactions between TRIM37 and PRC2 subunits could be demonstrated by co-immunoprecipitation (Fig. 3b and Extended Data Fig. 4a–c), liquid chromatography tandem mass spectroscopy (Extended Data Fig. 4d, e and Supplementary Table 2), an *in vitro* pull-down assay (Extended Data Fig. 4f), and the ability of a

GAL4–TRIM37 fusion protein to recruit PRC2 subunits to GAL4-binding sites (Extended Data Fig. 4g).

Global ChIP-chip analysis identified 3,384 genes that were co-bound by TRIM37 and EZH2 (Fig. 3c, Extended Data Fig. 2a and Supplementary Table 3). In addition, 3,560 genes were bound by TRIM37 and enriched for histone H3 lysine 27 trimethylation (H3K27me3) (Fig. 3d, Extended Data Fig. 2a, c and Supplementary Table 4). Knockdown of TRIM37 resulted in loss or reduction of H3K27me3 in ~65% of these genes (Fig. 3e, Extended Data Fig. 2a, d and Supplementary Table 4). Several studies have reported global correlations between PRC1/PRC2 occupancy and CpG islands^{14–18}. We found no statistically significant difference between the percentage of promoters co-bound by TRIM37 and EZH2 or bound by EZH2 alone that contain a CpG island (Extended Data Fig. 2e).

To validate the ChIP-chip results, we analysed a representative set of TRIM37 target genes for binding of EZH2 and BMI1 by directed ChIP assays. All gene promoters analysed were enriched for BMI1 and EZH2, which was diminished after TRIM37 knockdown (Fig. 3f and Extended Data Fig. 5a–c). These gene promoters were also enriched for H3K27me3, which was reduced after TRIM37 knockdown (Fig. 3g and Extended Data Fig. 5d). As expected, knockdown of BMI1 or EZH2 resulted in increased expression of these genes (Fig. 3h and Extended Data Fig. 5e, f).

Our results raised the possibility that in breast cancers containing amplified 17q23, TRIM37 is overexpressed and contributes to transformation by silencing tumour suppressor genes. To investigate the relationship

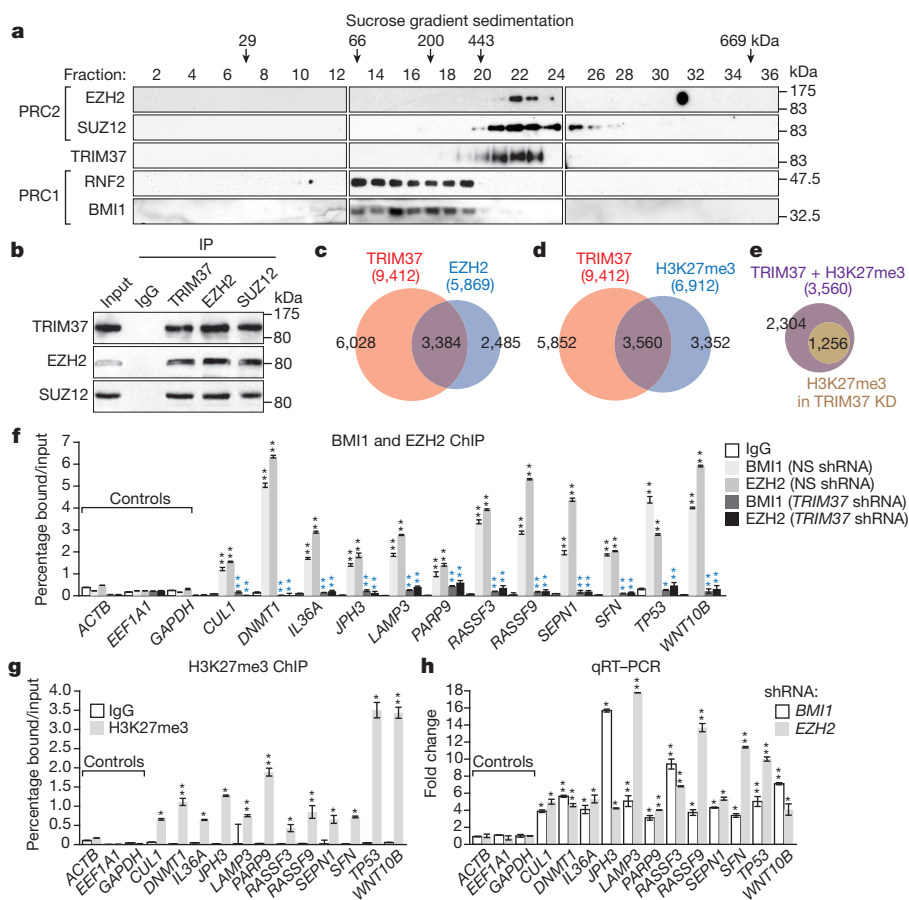


Figure 3 | Interaction and co-occupancy of TRIM37 and PRC2.

a, Immunoblots on sucrose gradient fractions. **b**, Sucrose gradient fractions 20–22 immunoprecipitated (IP) with a TRIM37, EZH2 or SUZ12 antibody, or an IgG control, were analysed by immunoblotting. **c–e**, Venn diagrams showing overlap between TRIM37- and EZH2-bound genes (**c**) and TRIM37-bound and H3K27me3-enriched genes (**d**) in parental MCF7 cells, or after knockdown of TRIM37 (**e**). See also Supplementary Tables 3 and 4. **f**, BMI1 and EZH2 ChIP in MCF7 cells expressing an NS or TRIM37 shRNA.

Error bars indicate s.d.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **g**, H3K27me3 ChIP in MCF7 cells. Error bars indicate s.d.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **h**, qRT-PCR monitoring TRIM37 target gene expression in BMI1- or EZH2-knockdown MCF7 cells. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). * $P < 0.05$; ** $P < 0.01$.

between TRIM37 levels and the expression of TRIM37 target genes, we analysed a published comprehensive data set of 466 human breast cancer samples¹⁹. Of the 9,412 TRIM37 target genes we identified, there was a statistically significant correlation between increased TRIM37 levels and decreased expression of 2,794 genes (Supplementary Table 5). The heat map shown in Fig. 4a illustrates the results for 60 TRIM37 target genes in a subset of breast cancer samples in which TRIM37 is either under- or overexpressed (Extended Data Fig. 6).

To confirm that elevated TRIM37 levels were responsible for decreased TRIM37 target gene expression, we analysed expression of representative

TRIM37 target genes in MCF10A cells ectopically expressing TRIM37. Expression of all TRIM37 target genes analysed was significantly lower in cells ectopically expressing TRIM37 compared with empty vector (Fig. 4b).

To investigate directly the potential oncogenic activity of TRIM37, we knocked down TRIM37 in MCF7 cells and measured the effect on tumorigenicity in mouse xenografts. Knockdown of TRIM37 markedly reduced tumour growth (Fig. 4c) without affecting proliferation of cultured MCF7 cells (Fig. 4d and Extended Data Fig. 7a). Knockdown of TRIM37 also reduced tumour formation in xenografts derived from BT474 and FC-IBC02 breast cancer cells (Extended Data Fig. 7b, c).

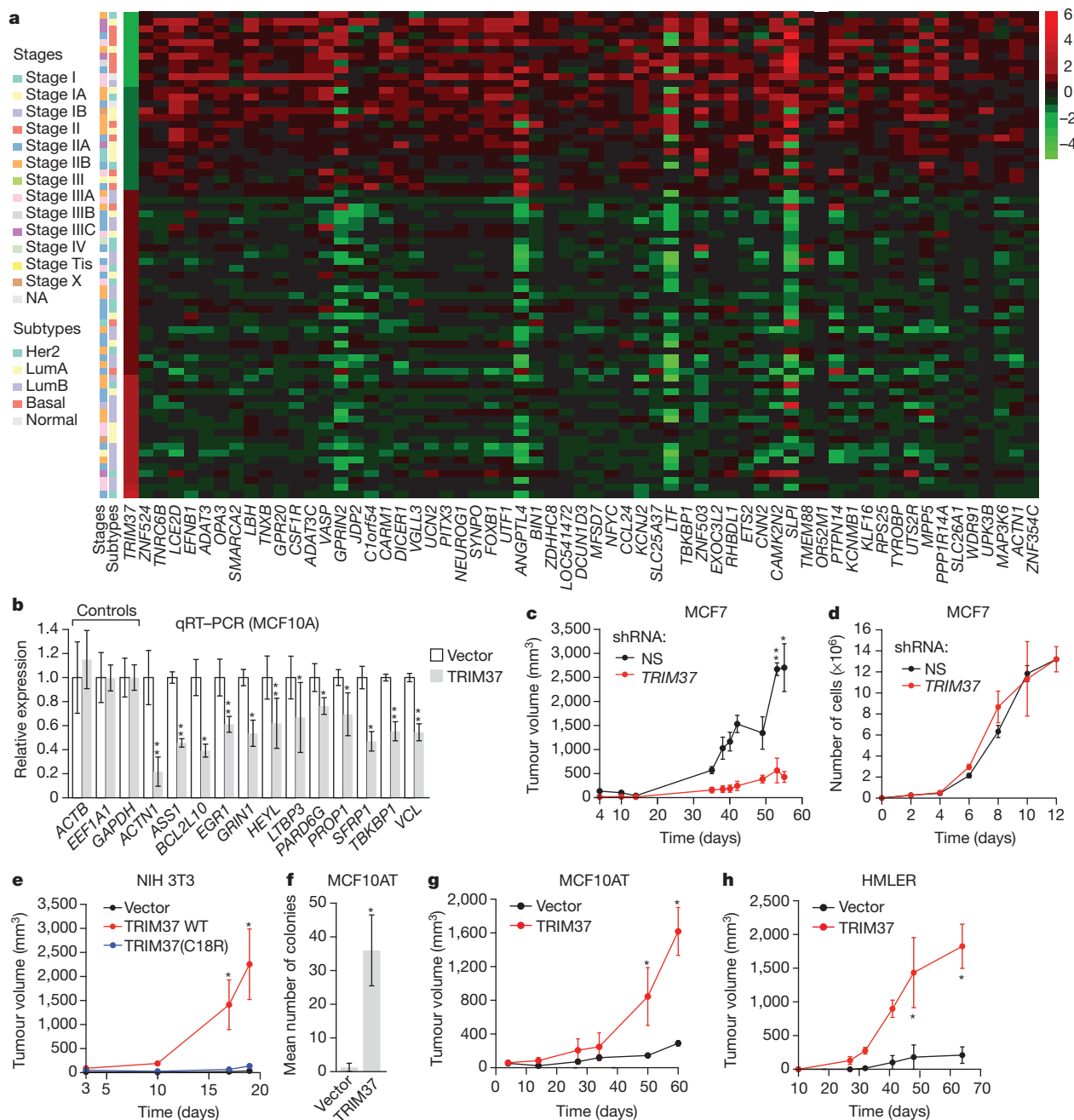


Figure 4 | TRIM37 is an oncogene. **a**, Heat map showing expression of TRIM37 and target genes in human breast cancer samples. See also Supplementary Table 5. **b**, qRT-PCR monitoring TRIM37 target gene expression in MCF10A cells expressing TRIM37 or vector. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **c**, Tumour formation in mice injected with MCF7 cells expressing an NS or TRIM37 shRNA. Error bars indicate s.e.m.; $n = 3$ mice per group. **d**, Proliferation of MCF7 cells expressing an NS or TRIM37 shRNA. Error bars indicate s.d.; $n = 3$

technical replicates of a representative experiment (out of three experiments). **e**, Tumour formation in mice injected with NIH 3T3 cells expressing vector, TRIM37 wild type (WT) or TRIM37(C18R). Error bars indicate s.e.m.; $n = 3$ mice per group. **f**, Soft agar assay with MCF10AT cells expressing vector or TRIM37. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **g**, **h**, Tumour formation in mice injected with MCF10AT (**g**) or HMLER (**h**) cells expressing vector or TRIM37. Error bars indicate s.e.m.; $n = 3$ mice per group. * $P < 0.05$; ** $P < 0.01$.

Conversely, ectopic expression of wild-type TRIM37 (Extended Data Fig. 7d), but not TRIM37(C18R), transformed NIH 3T3 cells, enabling tumour formation in mice (Fig. 4e). Moreover, ectopic expression of TRIM37 (Extended Data Fig. 7e) also transformed MCF10AT cells, a pre-malignant MCF10A derivative stably expressing activated *HRAS*²⁰, enabling growth in soft agar (Fig. 4f) and tumour formation in mice (Fig. 4g). Furthermore, knockdown of a subset of TRIM37 target genes with predicted tumour suppressor activity also transformed MCF10AT cells (Extended Data Fig. 7f–h). By contrast, TRIM37 could not transform parental MCF10A cells (Extended Data Fig. 7i, j), indicating that TRIM37 can cooperate with RAS to transform human breast epithelial cells. Ectopic expression of TRIM37 also transformed HMLER and BPLER cells, immortalized human breast epithelial cell lines containing activated *HRAS*²¹ (Fig. 4h and Extended Data Fig. 7k). Notably, the RAS pathway is activated in ~50% of breast cancers²².

We have identified a new H2A ubiquitin ligase and shown how it functions in conjunction with PRCs to silence target genes. In the canonical silencing pathway, PRC2 first interacts with the promoter, followed by EZH2-catalysed H3K27 trimethylation, which is recognized by PRC1, followed by RNF2-catalysed H2A ubiquitination^{8,13}. In cells containing high TRIM37 levels, TRIM37 associates with PRC2 and the TRIM37–PRC2 complex targets specific promoters. The promoter-bound PRC2–TRIM37 complex then carries out both H3K27 trimethylation and H2A mono-ubiquitination (see Extended Data Fig. 8a). Our results show that in the absence of PRC2, TRIM37 can catalyse H2A mono-ubiquitination (Fig. 1a, b) and bind to DNA (Fig. 3c). However, it remains possible that association with PRC2 could modulate, positively or negatively, the DNA-binding or H2A mono-ubiquitination activity of TRIM37. TRIM37 is also required for PRC1 occupancy, which presumably is mediated by PRC2 as in the canonical pathway. Although PRC1 does not direct H2A ubiquitination, it is still required for silencing, consistent with previous studies showing that PRC1 mediates other repressive activities, such as chromatin compaction^{23,24}.

TRIM37 alters the specificity of PRC2 and thus when TRIM37 is overexpressed there are extensive changes in gene expression that include the silencing of multiple tumour suppressor genes. Our results are consistent with previous studies showing that altered PRC2 activity, resulting, for example, from overexpression of EZH2, leads to malignancy^{25,26}. In support of our cell culture results, we found a correlation between increased TRIM37 levels and decreased expression of TRIM37 target genes in human breast cancer samples (Fig. 4a and Supplementary Table 5). Interestingly, there is also a statistically significant correlation between high levels of TRIM37 and decreased survival in patients with oestrogen-receptor-positive breast cancer (Extended Data Fig. 8b). Collectively, our results reveal TRIM37 as a new H2A ubiquitin ligase that is overexpressed and promotes transformation in a subset of breast cancers.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 October 2013; accepted 9 October 2014.

Published online 24 November 2014.

1. Sinclair, C. S., Rowley, M., Naderi, A. & Couch, F. J. The 17q23 amplicon and breast cancer. *Breast Cancer Res. Treat.* **78**, 313–322 (2003).
2. Budhidarmo, R., Nakatani, Y. & Day, C. L. RINGs hold the key to ubiquitin transfer. *Trends Biochem. Sci.* **37**, 58–65 (2012).
3. Weake, V. M. & Workman, J. L. Histone ubiquitination: triggering gene activity. *Mol. Cell* **29**, 653–663 (2008).
4. Wang, H. et al. Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873–878 (2004).

5. de Napoles, M. et al. Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev. Cell* **7**, 663–676 (2004).
6. Gazin, C., Wajapeyee, N., Gobeil, S., Virbasius, C. M. & Green, M. R. An elaborate pathway required for Ras-mediated epigenetic silencing. *Nature* **449**, 1073–1077 (2007).
7. Cao, R., Tsukada, Y. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol. Cell* **20**, 845–854 (2005).
8. Lanzuolo, C. & Orlando, V. Memories from the polycomb group proteins. *Annu. Rev. Genet.* **46**, 561–589 (2012).
9. Buchwald, G. et al. Structure and E3-ligase activity of the Ring–Ring complex of polycomb proteins Bmi1 and Ring1b. *EMBO J.* **25**, 2465–2474 (2006).
10. Richly, H. et al. Transcriptional activation of polycomb-repressed genes by ZRF1. *Nature* **468**, 1124–1128 (2010).
11. Monni, O. et al. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl Acad. Sci. USA* **98**, 5711–5716 (2001).
12. Kirmizis, A. et al. Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. *Genes Dev.* **18**, 1592–1605 (2004).
13. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).
14. Ku, M. et al. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **4**, e1000242 (2008).
15. Mohn, F. et al. Lineage-specific polycomb targets and *de novo* DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30**, 755–766 (2008).
16. Velichutina, I. et al. EZH2-mediated epigenetic silencing in germinal center B cells contributes to proliferation and lymphomagenesis. *Blood* **116**, 5247–5255 (2010).
17. Orlando, D. A., Guenther, M. G., Frampton, G. M. & Young, R. A. CpG island structure and trithorax/polycomb chromatin domains in human cells. *Genomics* **100**, 320–326 (2012).
18. Riising, E. M. et al. Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol. Cell* **55**, 347–360 (2014).
19. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
20. Miller, F. R. et al. Xenograft model of progressive human proliferative breast disease. *J. Natl. Cancer Inst.* **85**, 1725–1732 (1993).
21. Ince, T. A. et al. Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes. *Cancer Cell* **12**, 160–170 (2007).
22. von Lintig, F. C. et al. Ras activation in human breast cancer. *Breast Cancer Res. Treat.* **62**, 51–62 (2000).
23. Grau, D. J. et al. Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev.* **25**, 2210–2221 (2011).
24. Eskeland, R. et al. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* **38**, 452–464 (2010).
25. Chang, C. J. & Hung, M. C. The role of EZH2 in tumour progression. *Br. J. Cancer* **106**, 243–247 (2012).
26. Deb, G., Thakur, V. S. & Gupta, S. Multifaceted role of EZH2 in breast and prostate tumorigenesis: epigenetics and beyond. *Epigenetics* **8**, 464–476 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Peterson for providing *Xenopus* nucleosomes; R. Weinberg and M. Cristofanilli for providing cell lines; A. Virbasius and the University of Massachusetts Medical School (UMMS) RNAi Core Facility for providing shRNAs; P. Spatrick at the UMMS Genomics Core Facility; the UMMS Proteomics and Mass Spectroscopy Facility for mass spectrometry analysis; and S. Deibler for providing editorial assistance. N.W. is a Sidney Kimmel Scholar for Cancer Research and is supported by young investigator awards from the National Lung Cancer Partnership/Uniting Against Lung Cancer, Melanoma Research Alliance and International Association for the Study of Lung Cancer. This work was also supported by grants from CEA-DSV, ATIGE Genopole and Centre National de la Recherche Scientifique (CNRS) to C.G., and from the National Institutes of Health (R01GM033977) to M.R.G. M.R.G. is an investigator of the Howard Hughes Medical Institute.

Author Contributions C.G. made the initial observation that TRIM37 has an H2A mono-ubiquitination activity. S.B., C.G., J.S.T. and M.R.G. designed the experiments. S.B., L.C., X.Z., L.L. and N.W. performed the research. C.-M.V. provided critical reagents. J.O. and L.J.Z. performed biostatistical analysis for ChIP-chip experiments and database mining. S.B. and M.R.G. analysed and interpreted the data and wrote the paper. All authors reviewed the paper and provided comments.

Author Information The ChIP-chip data have been deposited in the Gene Expression Omnibus under accession number GSE48196. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.R.G. (michael.green@umassmed.edu).

METHODS

Cell lines and cell culture. K: Molv NIH 3T3 cells (ATCC# CRL-6361; referred to here as *Kras* NIH 3T3) were maintained in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal calf serum (FCS) at 37 °C and 5% CO₂. HMECs (Life Technologies) were cultured in HuMec Ready Media (Life Technologies). MCF10A (ATCC) and MCF10AT²⁷ cells were maintained in DMEM/F12 supplemented with 5% horse serum (Invitrogen), EGF (Peprotech), hydrocortisone (Sigma), cholera toxin (Sigma), insulin (GIBCO) and penicillin/streptomycin (Life Technologies). MCF7 cells (National Cancer Institute) were maintained in DMEM supplemented with 1 × nonessential amino acids (NEAA) and 10% fetal bovine serum (FBS). BT474 cells (ATCC) were maintained in ATCC HybriCare Media supplemented with 10% FBS. MDA-MB-361 and MDA-MB-231 cells (ATCC) were cultured in ATCC-formulated Leibovitz's L-15 medium supplemented with 20% FBS. T47D cells (National Cancer Institute (NCI)) were maintained in RPMI medium supplemented with 10% FBS. Hs578T cells (NCI) were maintained in DMEM medium supplemented with 0.01 mg ml⁻¹ bovine insulin and 10% FBS. BPLER and HMLER cells²¹ (provided by R. Weinberg, Whitehead Institute, Massachusetts Institute of Technology) were maintained in WIT (Stemgent) and mammary epithelial cell growth medium (Lonza), respectively. FC-IBC02 cells²⁸ (provided by M. Cristofanilli, Thomas Jefferson University) were maintained in Ham's F12 Nutrient Mixture (Life Technologies) with 10% FBS and 0.5% insulin (Gibco) with 100 µg l⁻¹ hydrocortisone (Sigma) and antibiotic-antimycotic (Gibco). Cells cultured at the same time were pooled together and then seeded after counting in a 6-well or 10-cm dish. Dishes/wells were then subjected, in a random order, to treatment with a control or test shRNA/vector.

RNA interference. For stable shRNA knockdowns, cells were seeded in a six-well plate to 50% confluency and subsequently transduced with 400 µl lentiviral particles expressing shRNAs (obtained from Open Biosystems/Thermo Scientific through the UMMS RNAi Core Facility, listed in Extended Data Table 1) in a total volume of 2 ml of appropriate media supplemented with 6 µg ml⁻¹ polybrene. Media was replaced after overnight incubation to remove the polybrene and viral particles and cells were subjected to puromycin selection (2 µg ml⁻¹) for 3 days.

In vitro ubiquitination assays. To produce purified recombinant TRIM37, the TRIM37 open reading frame was PCR amplified from a cDNA clone (Invitrogen) using high fidelity Pfu polymerase, and cloned into a derivative of pEF6/V-5HisB (Invitrogen), resulting in the addition of a protein-C epitope (MAEDQVDPRLID GKEFT) at the N terminus of TRIM37. The C18R mutation was introduced by overlap extension PCR using primers listed in Extended Data Table 2. All constructs were verified by sequencing.

In vitro ubiquitination reactions were set up containing purified TRIM37 or TRIM37(C18R) (500 ng to 2 µg), a combination of BMI1 (500 ng; Abcam) and RNF2 (500 ng; Origene) mixed together, or BRCA1 (500 ng; Abcam) plus H2A or H2B (5X; Boston Biochemicals), 10 nM UBE1 (E1; Boston Biochemicals), 10 nM UBCH5B (E2; Boston Biochemicals), and 10 nM HA-ubiquitin aldehyde (Boston Biochemicals) in 10 µl of reaction buffer (50 mM Tris-Cl, pH 7.9, 5 mM MgCl₂, 2 mM NaF, 100 mM dithiothreitol (DTT) and 2 mM ATP). Reactions were incubated at 32 °C for 90 min. For the nucleosomal assay, *in vitro* ubiquitination reactions were set up similarly but contained *Xenopus* oocyte-derived nucleosomes; recombinant histone particles were expressed and purified as described previously²⁹ and reconstituted using a standard salt step dialysis method³⁰. To generate the H2A(KKRR) mutant¹⁰, a pET-based plasmid expressing H2A (provided by C. Peterson, University of Massachusetts Medical School) was used as a template in a PCR-based site-directed mutagenesis with primer extension reaction using primers listed in Extended Data Table 2. Proteins were resolved by 15% SDS-PAGE and subjected to immunoblotting with H2Aub(K119) (Cell Signaling Technology #8240, clone D27C4), H2A (Cell Signaling Technology #2578), H2Bub (EMD Millipore 05-1312, clone 56), H2B (Abcam ab18977), H3 (Cell Signaling Technology #9715) or H4 (Cell Signaling Technology #2592) antibodies.

qRT-PCR. Total RNA was isolated and reverse transcribed using Superscript II Reverse Transcriptase (Invitrogen). Quantitative real-time PCR was performed as described previously⁶ using primers listed in Extended Data Table 2. Gene expression was normalized to *GAPDH*. Controls lacking reverse transcriptase were carried out in parallel to rule out the possibility of DNA contamination. Each sample was analysed three independent times and the results from one representative experiment, with technical triplicates, are shown.

Immunoblotting. Nuclear lysates were prepared as previously described³¹. Nuclear extracts were separated on either 8%, 10% or 20% SDS-PAGE gels. Blots were probed with primary antibodies (H2Aub(K119), Cell Signaling Technology; H2A, Cell Signaling Technology; H2Bub, EMD Millipore; H2B, Cell Signaling Technology #12364, clone D2H6; H, Cell Signaling Technology; H4, Cell Signaling Technology; HA, Cell Signaling Technology #2367, clone 6E2; TRIM37, Abcam ab95997, or custom made by 21st Century Biochemicals against a synthetic peptide corresponding to amino acids 444–460 of the human protein followed by affinity purification;

RNF2, Abcam ab28629; and α -tubulin, created in house) overnight at 4 °C, washed five times in TBS plus 0.1% Tween (TBST) and then incubated with the appropriate HRP-conjugated secondary antibody for 1 h at room temperature. Membranes were washed five times in TBST and visualized on autoradiography film after incubating with ECL reagent (Supersignal West Pico or Supersignal West Femto; Thermo Scientific). Immunoblots were quantified using ImageJ software v.1.47 (National Institutes of Health).

TRIM37 overexpression. A TRIM37 complementary DNA clone (Origene Technologies) was subcloned into the vector p3X-Flag-Myc-CMV-26 (Sigma) using NotI and XbaI sites, and verified by full-length sequencing. The TRIM37 expression vector was transfected into cells using Effectene reagent (Qiagen) and stable clones were selected. Overexpression of TRIM37 was confirmed by immunoblotting using TRIM37 (21st Century) and anti-Flag (Sigma M2) antibodies.

ChIP-chip and directed ChIP. For ChIP-chip analysis, chromatin was sonicated and immunoprecipitation was carried out as previously described³¹ using a TRIM37 (Abcam), H2Aub (Cell Signaling Technology), EZH2 (Cell Signaling Technology #5246, clone D2C9), or H3K27me3 (Cell Signaling Technology #9733, clone C36B11) antibody. Samples were prepared and hybridized to a GeneChip Human Promoter 1.0R Array (Affymetrix) according to the manufacturer's specifications, and arrays were scanned using a GeneChip Scanner 3000 7G (Affymetrix). Each ChIP-chip experiment was performed in duplicate. ChIP-chip data were analysed using Bioconductor package Starr^{32,33} for data read-in, normalizing and peak calling. Quantile normalization was applied to normalize probe intensity before peak calling with default settings except that fragment length (sonication fragment length) was set to 500, minrun (minimum number of probes in a bound region) was set to 2, method (multiple test adjustment method) was set to BH (Benjamini & Hochberg)³⁴, and α (false discovery rate cut-off) was set to 0.1. ChIP-chip results were annotated by an R (v.3.1.0)/Bioconductor (v.2.14)/ChIPpeakAnno(v.2.12.1) package³⁵. Density plots were plotted by the distance of peaks to the annotated transcription start site. Venn diagrams were drawn by an R (v.3.1.0)/Vennable (v.3.0) package. Gene ontology analysis was done by the R/Bioconductor package goProfiles³⁶. For the statistical analysis on the global decrease in H3K27me3 and H2A-ub following TRIM37 knockdown in MCF7 cells, the Wilcoxon signed-rank test was performed in R³⁷. A summary table showing the overlap of proteins and epigenetic marks, along with the *P* values for each overlap, is shown in Supplementary Table 6. CpG island overlap analysis was performed as previously described¹⁷ except that CpG island annotation was downloaded from the UCSC Genome Browser (<https://www.genome.ucsc.edu>). A CpG gene list was generated for CpG islands located within 1 kb upstream of the transcription start site. The percentage of promoters bound by EZH2 alone or co-bound by TRIM37 and EZH2 and which also contain a CpG island was calculated by the intersection of the CpG gene list and the EZH2 or EZH2/TRIM37 target gene list, respectively. The ChIP-chip data have been deposited in NCBI's Gene Expression Omnibus³⁸ and are accessible through accession number GSE48196.

Directed ChIP assays were performed essentially as described previously⁶ using cell extracts prepared 7 days after retroviral transduction and puromycin selection, and antibodies against BMI1 (Abcam ab14389, clone 1.T.21), EZH2 (Cell Signaling Technology), H3K27me3 (Cell Signaling Technology), H2Aub (Cell Signaling Technology) and TRIM37 (Abcam) or, as a negative control, IgG. The sequences of primers used for amplifying ChIP products are listed in Extended Data Table 2. On the basis of the probe sequence used in the Affymetrix array, for each gene a 400–600 bp region was tested for enrichment of TRIM37 using at least three primer pairs. The primer pair with the consistently highest enrichment in three different experiments was chosen for subsequent enrichment analysis of BMI1, H2Aub, H3K27me3 and EZH2. Each ChIP experiment was carried out three independent times and the results from one representative experiment, with technical triplicates, are shown.

Identification of putative tumour suppressors. Candidate TRIM37 target genes were systematically queried against three databases: DNA Tumour Suppressor and Oncogene Database (DTSOD; available at <http://embryology.med.unsw.edu.au>), TSGene database (TSGene; <http://bioinfo.mc.vanderbilt.edu/TSGene/>), and Tumour Associated Gene (TAG; <http://www.binfo.ncu.edu.tw/TAG/GeneDoc.php>). Genes listed in the TSGene database or listed as "tumor suppressor genes" in the DTSOD or TAG database were considered putative tumour suppressor genes. Statistical enrichment of putative tumour suppressor genes among TRIM37 target genes was calculated by a hypergeometric test.

Sucrose gradient fractionation. Sucrose gradient sedimentation analysis was performed as described³⁹. Briefly, 10–40% gradients were formed by layering 500 µl NEB1 buffer containing 10%, 20%, 30% or 40% sucrose in a 11 × 34 mm centrifuge tube (Beckman) and allowed to equilibrate at room temperature for 2 h. Gradients were chilled, loaded with 500 µg MCF7 nuclear extract (adjusted to a volume of 150 µl) or 150 µl molecular weight markers (Sigma MW-GF-1000), and centrifuged in a Beckman TLS-55 rotor at 214,000g for 14 h. Thirty-six fractions of 45 µl were collected. For the markers, 20 µl of each fraction was electrophoresed and Coomassie

stained. For the gradient fractions, 20 µl of fractions were analysed by immunoblotting using EZH2 (Cell Signaling Technology), SUZ12 (Abcam ab12073), TRIM37 (Abcam), RNF2 (Abcam ab101273) and BMI1 (Bethyl Laboratories, A301-694A).

Co-immunoprecipitation. MCF7 nuclear extract (~1,000 µg) was incubated with a TRIM37 (Abcam), EZH2 (Cell Signaling Technology) or SUZ12 (Abcam) antibody at 4 °C for 24 h in the presence or absence of ethidium bromide (100 µg ml⁻¹). For immunoprecipitations from fractionated nuclear lysate, fractions 20–22 (which were enriched for TRIM37, EZH2 and SUZ12) were pooled and diluted 15-fold. Immune complexes were captured on rabbit or mouse TrueBlot IP beads (eBioScience), washed three times in NEB1 buffer, and eluted by boiling for 10 min in 2× SDS sample buffer. Immunoprecipitations from MCF7 cells ectopically expressing Flag-tagged TRIM37 were performed similarly, except that anti-Flag magnetic beads (Sigma) were used to capture the immune complexes. Immunoprecipitated proteins were analysed by immunoblotting as described earlier. Input lanes represent 10–25% of extract loaded in the immunoprecipitation lanes.

Liquid chromatography tandem mass spectrometry. Large-scale immunoprecipitation was carried out as previously described⁴⁰ by incubating 80 mg MCF7 nuclear extract with 50 µg of anti-TRIM37 antibody (Abcam) at 4 °C for 5 h. Immune complexes were captured on IgG Dynabeads (Life Technologies) by additional incubation at 4 °C for 2 h. The beads were then washed with NETN (20 mM Tris-HCl (pH 8.0), 0.1 M NaCl, 1 mM EDTA, 0.5% NP-40) three times and re-suspended in Laemmli buffer. Proteins were separated on a short 10% SDS-PAGE. A 2 cm gel slice with all the proteins was submitted for mass spectrometry at the University of Massachusetts Medical School Proteomics and Mass Spectrometry Facility.

For in-gel digestion, gel slices were placed in 1.5 ml Eppendorf tubes with 1 ml of water for 30 min and then replaced with 100 µl of 250 mM ammonium bicarbonate. For reduction, 20 µl of 45 mM DTT was added and the samples were incubated at 50 °C for 30 min. The alkylation was carried out by addition of 20 µl of a 100 mM iodoacetamide solution for 30 min. The gel slices were washed twice with water and then treated with 50:50 50 mM ammonium bicarbonate:acetonitrile for 1 h. The solution was then removed and 200 µl of acetonitrile was added. Dried gel slices were rehydrated in 50 µl of 2 ng µl⁻¹ trypsin (Sigma) in 0.01% ProteaseMAX Surfactant (Promega):50 mM ammonium bicarbonate at 37 °C for 21 h. Additional dehydration was carried out with 100 µl of 80:20 acetonitrile:1% formic acid. Dried samples were resuspended in 25 µl of 5% acetonitrile in 0.1% trifluoroacetic acid.

Samples were injected onto a custom packed 2 cm × 100 µm C₁₈ Magic 5 µm particle trap column. Labelled peptides were then eluted and sprayed from a custom packed emitter (75 µm × 25 cm C₁₈ Magic 3 µm particle) with a linear gradient from 95% solvent A (0.1% formic acid in water) to 35% solvent B (0.1% formic acid in acetonitrile) in 90 min at a flow rate of 300 nl per minute on a Waters Nano Acquity UPLC system. Data-dependent acquisitions were performed on a Q Exactive mass spectrometer (Thermo Scientific). Raw data files were processed with Proteome Discoverer (v.1.3) before searching with Mascot Server (v.2.4) against the Human Index of the SwissProt database. Search parameters used were fully tryptic with two missed cleavages, parent mass tolerances of 10 p.p.m. and fragment mass tolerances of 0.05 Da. A fixed modification of carbamidomethyl cysteine and variable modifications of acetyl (protein N-terminus), pyro glutamic for N-terminal glutamine, oxidation of methionine was considered. Search results were loaded into the Scaffold Viewer (Proteome Software), which provided total spectra scores and interaction probabilities.

GST pull-down assay. The TRIM37 cDNA (Origene Technologies) was cloned into the vector pGEX4T-1 (Life Technologies) using EcoRI and NotI sites and the construct was verified by full-length sequencing. Supernatant from bacteria expressing GST-TRIM37 was incubated with glutathione agarose beads (GE Healthcare) at 4 °C for 3–4 h. Beads were washed three times with PBS and 0.5% Triton X-100 before using in pull-down assays. PRC2 subunits (EZH2 (Origene Technologies), SUZ12 (Origene Technologies), AEBP2 (Open Biosystems), EED (Open Biosystems) and RBAP48 (Open Biosystems)) were *in vitro* translated using the TnT T7 Quick Coupled Transcription/Translation System (Promega) according to the manufacturer's specifications. GST-TRIM37 beads (3 µg) were incubated with *in vitro* translated protein (added to 1/10 volume of the reaction mixture) in Buffer A (50 mM Tris-HCl pH 7.9, 0.5 mM EDTA, 1 mM DTT, 0.2 mM PMSF, 10% glycerol, 150 mM KCl and 0.05% NP40) at 4 °C for 2 h. After incubation, beads were washed twice with wash buffer (50 mM Tris-HCl pH 7.9, 0.5 mM EDTA, 1 mM DTT, 0.2 mM PMSF, 10% glycerol, 300 mM KCl and 0.05% NP40). The interaction was detected by immunoblot using a biotin antibody (Abcam #53494).

GAL4-TRIM37 recruitment assay. The TRIM37 cDNA was cloned into the plasmid pCMX-Gal4 (Addgene), which contains amino acids 1–147 of GAL4, using KpnI and NheI sites. TRIM37 deletions were generated using PCR-based site-directed deletion with primer extension using primers listed in Extended Data Table 2. All constructs were verified by full-length sequencing. 293T cells were co-transfected with the GAL4-TRIM37 plasmid and G5E1bcat-pSP72 (ref. 41) (containing five

GAL4-binding sites) or E1bcat-pSP72 (ref. 42) (negative control). ChIP assays were carried out as described earlier.

Breast cancer bio-statistical analysis. Gene expression and CpG methylation data from breast cancer samples ($n = 466$) were obtained from The Cancer Genome Atlas (TCGA) Data Portal (https://tcga-data.nci.nih.gov/docs/publications/brca_2012/)¹⁹. The mRNA data were converted to relative mRNA expression levels by subtracting each gene's median level computed over 22 tumour-adjacent normal controls from its level observed in each patient sample. A subset of samples was then selected for further analysis to include those with upregulated TRIM37 ($n = 45$ with TRIM37 value ≥ 1) and downregulated TRIM37 ($n = 26$ with TRIM37 value ≤ -1). Pearson correlation analysis was performed to identify TRIM37 target genes (identified in the ChIP-chip experiment) whose downregulation significantly correlated with increased TRIM37 expression ($P < 0.01$). Statistical analysis and plotting were done with R (v.3.1.0), a system for statistical computation and graphics³⁷. The top 60 genes with negative correlation to TRIM37 were extracted for heat map plots using an R (v.3.1.0)/pheatmap (v.0.7.7) package. Cancer subtypes were extracted from sample list information of gene expression data (level 3, PAM50). Cancer stages were extracted from TCGA clinical information.

TRIM37/17q23 copy number data were also downloaded from https://tcga-data.nci.nih.gov/docs/publications/brca_2012/, section "SNP and Copy Number" Level 3 Data Archives. Data were visualized using R/Bioconductor package Gviz and analysed as described previously¹⁹. Briefly, thresholded gene level copy number values were calculated using GISTIC 2.0 and log₂ copy number values were thresholded as followed: values < -1 are set to -2 , values between -1 and -0.3 are set to -1 , values between -0.3 and 0.3 are set to 0 , values between 0.3 and 1 are set to 1 , and values > 1 are set to 2 . Tumours in which all gene level thresholded copy number changes equalled 0 were excluded from the analysis.

For correlation analysis between TRIM37 and prognosis, Kaplan–Meier plotting and log-rank test were done at <http://kmplot.com/analysis/index.php?p=service&cancer=breast>. Analysis was restricted to breast cancer patients with ER-positive subtypes. Patients were divided into two classes based on TRIM37 expression: patients with higher than upper quartile TRIM37 expression were considered to belong to the high-expressing TRIM37 group ($n = 352$) and the other patients were considered to belong to the low-expressing TRIM37 group ($n = 1,061$).

Animal experiments. All animal protocols were approved by the Institution Animal Care and Use Committee at UMMS (A-2016). Animal sample sizes were selected based on precedents established from previous publications and an understanding that at least $n = 5$ is generally required to achieve statistical significance. For mouse experiments involving shRNAs, the most efficacious shRNA of multiple shRNAs tested and validated in cell culture was used, a criterion that was established prospectively. Mice were randomly allocated to each group. No blinding was done as animal groups were identified by tagging and labelling the cages with the cells injected. Animals were excluded, according to pre-established criteria, if the tumour volume reached $> 2,000 \text{ mm}^3$; if tumour size or location affected the mobility or general health of animal, the animal was euthanized and excluded from the experiment or the complete experiment was terminated.

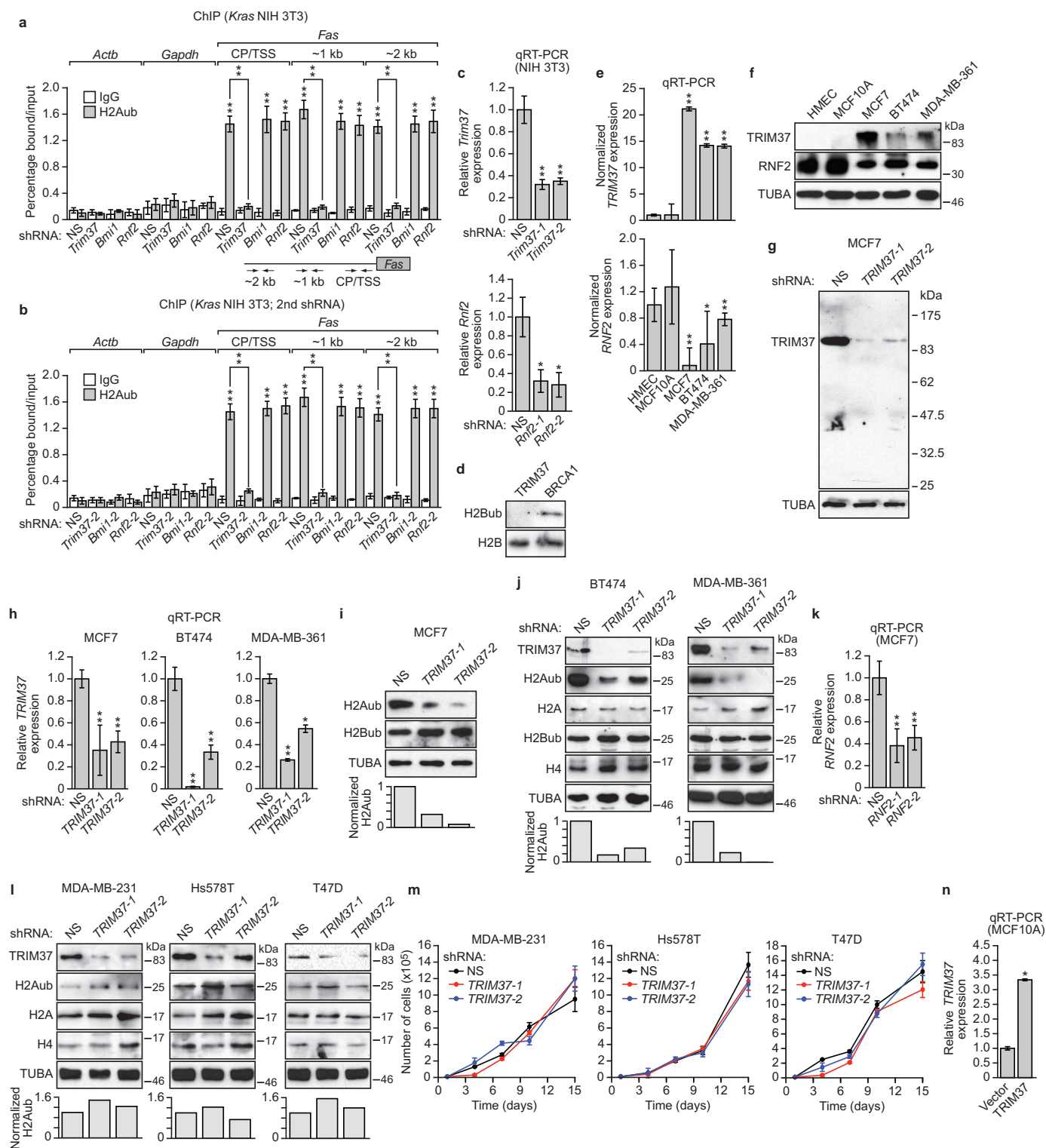
MCF7 (1×10^6), BT474 (3×10^6) or FC-IBC02 (3×10^6) cells stably expressing a non-silencing (NS) or TRIM37 shRNA were injected into the right thoracic mammary fat pad (MCF7) or flank (BT474 and FC-IBC02) of 6–8-week-old BALB/c nu/nu female mice ($n = 3$; Taconic Farms) to produce orthotopic primary tumours. Because MCF7, BT474 and FC-IBC02 cells are hormone dependent, a 60 day release tablet of 1.7 mg oestradiol (Innovative Research of America) was implemented under the skin of the neck of the mouse with a 11-gauge trocar. Primary tumour growth was evaluated by measurement with callipers. NIH 3T3 (5×10^6), MCF10AT (3×10^6), HMLER (3×10^6) or BPLER (3×10^6) cells expressing vector or TRIM37 or TRIM37(C18R) were subcutaneously injected in the flank of the mice ($n = 3$) and the growth of the primary tumours was monitored as previously described⁶. MCF10AT (5×10^6) cells expressing an NS shRNA or shRNA targeting a tumour suppressor gene were subcutaneously injected in the flank of the mice ($n = 3$) and tumour growth was monitored. All mouse experiments were repeated at least twice (for a total of at least 6 mice per group).

Soft agar assay. MCF10AT cells (2.5×10^4) expressing vector or TRIM37, or MCF10AT cells expressing an NS shRNA or shRNA targeting a tumour suppressor gene, were re-suspended in 3 ml of top agar (MCF10AT media containing 0.3% Difco Noble agar (BD Biosciences)) warmed to 40 °C. The cell suspension was layered onto 3 ml of set bottom agar (MCF10AT media containing 0.7% Noble agar) in a 6-well plate. Two millilitres of medium was added on the top agar the next day and changed every other day. Visible colonies were scored after 4–5 weeks.

Statistics. To achieve statistical significance, all qRT-PCR and ChIP data were collected from experiments performed in technical triplicate; each experiment was repeated at least twice and statistically significant results were obtained in independent biological replicates. Differences between groups were assayed using a two-tailed student *t*-test using Microsoft Excel. In cases where the assumption of the

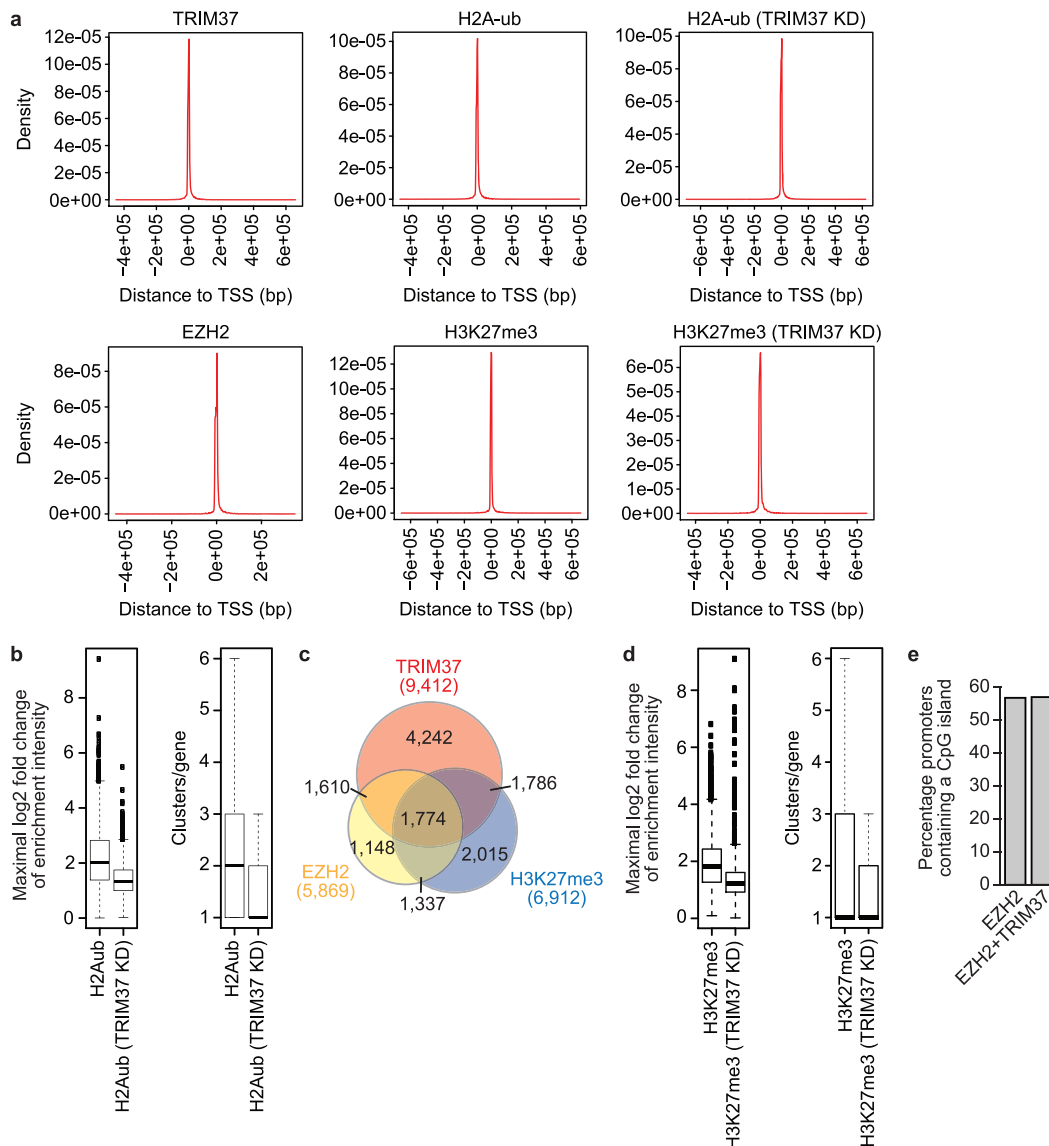
t-test was not valid, a non-parametric statistical method was used (for example, Wilcoxon signed-rank test). Significant differences were considered when $P < 0.05$; $*P \leq 0.05$ and $**P \leq 0.01$. Error bars indicate the standard deviation or standard error of the mean for the technical replicates, as indicated in figure legends.

27. Dawson, P. J., Wolman, S. R., Tait, L., Heppner, G. H. & Miller, F. R. MCF10AT: a model for the evolution of cancer from proliferative breast disease. *Am. J. Pathol.* **148**, 313–319 (1996).
28. Fernandez, S. V. *et al.* Inflammatory breast cancer (IBC): clues for targeted therapies. *Breast Cancer Res. Treat.* **140**, 23–33 (2013).
29. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol. Biol.* **119**, 1–16 (1999).
30. Hansen, J. C., Ausio, J., Stanik, V. H. & van Holde, K. E. Homogeneous reconstituted oligonucleosomes, evidence for salt-dependent folding in the absence of histone H1. *Biochemistry* **28**, 9129–9136 (1989).
31. Maston, G. A. *et al.* Non-canonical TAF complexes regulate active promoters in human embryonic stem cells. *eLife* **1**, e00068 (2012).
32. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
33. Zacher, B., Kuan, P. F. & Tresch, A. Starr: Simple Tiling ARray analysis of Affymetrix ChIP-chip data. *BMC Bioinformatics* **11**, 194 (2010).
34. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* **57**, 289–300 (1995).
35. Zhu, L. J. *et al.* ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).
36. Salicrú, M., Ocana, J. & Sanchez-Pla, A. Comparison of lists of genes based on functional profiles. *BMC Bioinformatics* **12**, 401 (2011).
37. Ihaka, R. & Gentleman, R. R. a language for data analysis and graphics. *J. Comput. Graph. Stat.* **5**, 299–314 (1996).
38. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
39. Tanese, N. Small-scale density gradient sedimentation to separate and analyze multiprotein complexes. *Methods* **12**, 224–234 (1997).
40. Wang, Y. *et al.* BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev.* **14**, 927–939 (2000).
41. Liu, F. & Green, M. R. A specific member of the ATF transcription factor family can mediate transcription activation by the adenovirus E1a protein. *Cell* **61**, 1217–1224 (1990).
42. Lillie, J. W. & Green, M. R. Transcription activation by the adenovirus E1a protein. *Nature* **338**, 39–44 (1989).



Extended Data Figure 1 | Control experiments for Fig. 1. **a**, ChIP monitoring H2Aub enrichment on *Fas* in *Kras* NIH 3T3 cells expressing a non-silencing (NS) shRNA or a *Trim37*, *Rnf2* or *Bmi1* shRNA. Three regions of the *Fas* promoter were analysed: the core promoter/transcription start site (CP/TSS), and 1 and 2 kb upstream of the TSS. *Actb* and *Gapdh* are shown as negative controls. Error bars indicate s.e.m.; $n = 9$ (three biological replicates with three technical replicates per sample). **b**, H2Aub ChIP as described in **a** using a second *Trim37*, *Rnf2* or *Bmi1* shRNA unrelated to that used in **a**. **c**, qRT-PCR analysis monitoring knockdown efficiencies of *Trim37* and *Rnf2* shRNAs in NIH 3T3 cells. The results are given relative to expression after treatment with an NS shRNA, which was set to 1. For knockdown efficiencies of the *Bmi1* shRNAs see ref. 6. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **d**, *In vitro* ubiquitination assay. Purified H2B was incubated with E1 (UBE1), E2 (UBCH5B), E3 (TRIM37 or BRCA1), ATP and HA-tagged ubiquitin. Blots were probed with antibodies against H2Bub or H2B. The results show that TRIM37 does not ubiquitinate H2B. BRCA1, which is known to ubiquitinate H2B at K120 (refs 39, 40) was used as a positive control. **e**, **f**, qRT-PCR (**e**) and immunoblots (**f**) monitoring TRIM37 and RNF2 in various cell lines. Expression of TRIM37 and RNF2 was normalized to that obtained in HMECs, which was set to 1. Error bars indicate s.d.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **g**, Immunoblot monitoring TRIM37 levels in MCF7 cells expressing an NS or one of two unrelated *TRIM37* shRNAs. α -Tubulin (TUBA) was monitored as a loading control. The results indicate that the TRIM37 antibody is highly specific. **h**, qRT-PCR analysis monitoring knockdown efficiencies of *TRIM37* shRNAs

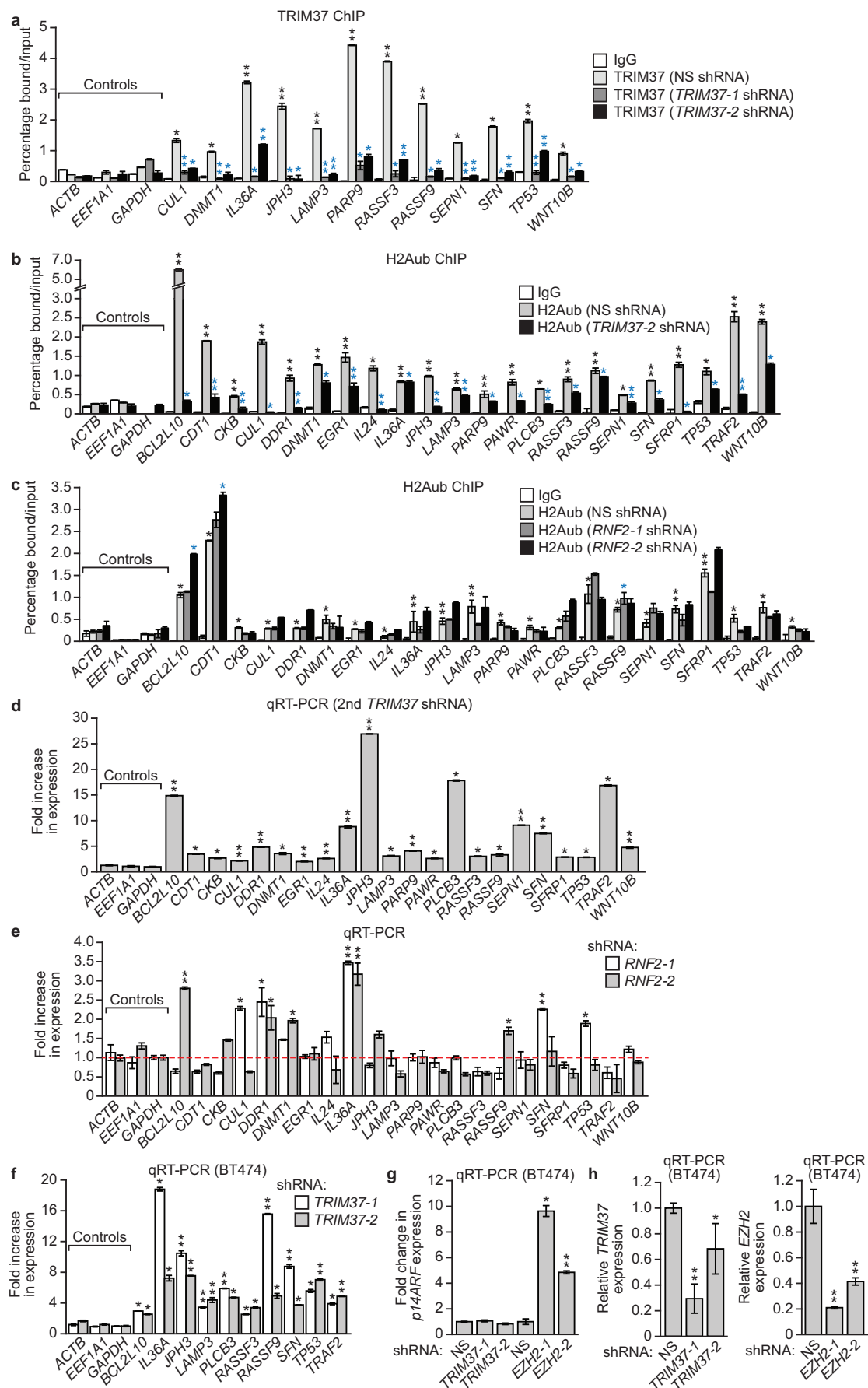
in MCF7, BT474 and MDA-MD-361 cells. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **i**, Top, immunoblots monitoring levels of H2Aub and H2Bub in MCF7 cells expressing an NS or one of two unrelated *TRIM37* shRNAs. Bottom, quantification of the H2Aub immunoblots relative to TUBA. The relative level of H2Aub in NS cells was set to 1. In this experiment, histones were acid extracted. **j**, Top, immunoblots monitoring TRIM37, H2Aub, H2A, H2Bub and H4 in BT474 and MDA-MB-361 cells expressing an NS or one of two unrelated *TRIM37* shRNAs. Bottom, H2Aub quantification, as described in **g**. **k**, qRT-PCR analysis monitoring knockdown efficiencies of *RNF2* shRNAs in MCF7 cells. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **l**, Top, immunoblots monitoring TRIM37, H2Aub, H2A and H4 in MDA-MB-231, Hs578T and T47D cells expressing an NS or one of two unrelated *TRIM37* shRNAs. Bottom, H2Aub quantification, as described in **g**. **m**, Proliferation of cultured MDA-MB-231, Hs578T and T47D cells expressing an NS or *TRIM37* shRNA. Error bars indicate s.d.; $n = 3$ technical replicates of a representative experiment (out of three experiments). The results show that knockdown of TRIM37 has no effect on proliferation of breast cancer cell lines lacking 17q23 amplification. **n**, qRT-PCR analysis monitoring *TRIM37* expression in an MCF10A cell line ectopically expressing TRIM37 or, as a control, empty vector. The results were normalized to TRIM37 expression in MCF10A cells expressing empty vector, which was set to 1. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). * $P < 0.05$, ** $P < 0.01$.



Extended Data Figure 2 | Extra details related to ChIP-chip analysis.

a, Histograms showing enrichment of TRIM37, H2Aub (in parental MCF7 cells or after TRIM37 knockdown (KD)), EZH2 and H3K27me3 (in parental MCF7 cells or after TRIM37 knockdown) as a function of distance to the nearest transcription start site (TSS). **b**, Left, box plot of maximal gene-level \log_2 fold change of enrichment intensity comparing input for H2Aub in parental MCF7 cells and after TRIM37 knockdown. Right, box plot of number of enriched regions per gene for H2Aub in parental MCF7 cells or after TRIM37 knockdown. The results show that there was a significant difference in H2Aub enrichment at TRIM37 target genes after TRIM37 knockdown with respect to both enrichment intensity ($P < 1 \times 10^{-22}$) and number of enriched regions ($P = 2.787008 \times 10^{-222}$). A region with a fold change ≥ 2 and a false discovery rate < 0.1 was considered a differentially enriched site. **c**, Venn diagram showing overlap between TRIM37- and EZH2-bound genes and H3K27me3-enriched genes in MCF7 cells. **d**, Left, box plot of maximal gene-level \log_2 fold

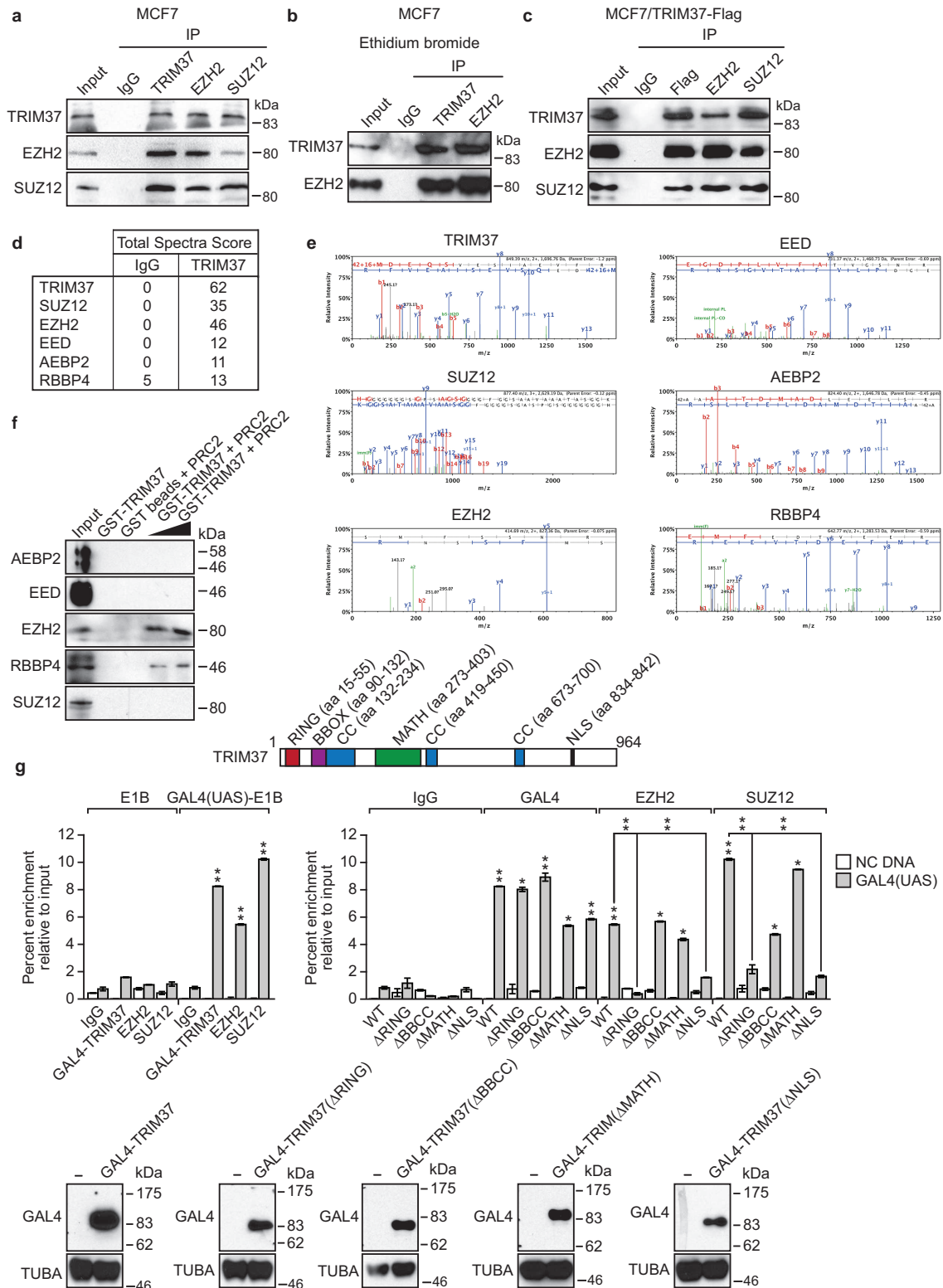
change of enrichment intensity comparing input for H3K27me3 in parental MCF7 cells or after TRIM37 knockdown. Right, box plot of number of enriched regions per gene for H3K27me3 in parental MCF7 cells or after TRIM37 knockdown. The results show that there was a significant difference in H3K27me3 enrichment at TRIM37 target genes after TRIM37 knockdown with respect to both enrichment intensity ($P = 5.872777 \times 10^{-109}$) and number of enriched regions ($P = 1.178392 \times 10^{-52}$). A region with a fold change ≥ 2 and a false discovery rate < 0.1 was considered a differentially enriched site. **e**, Percentage of promoters (defined as 1 kb upstream of the TSS) bound by EZH2 alone that contain a CpG island (3,332/5,869; 56.77%) or co-bound by EZH2 and TRIM37 that contain a CpG island (1,929/3,384; 57.00%). A two-sample test for equality of proportions with continuity correction showed that there was no statistically significant difference in CpG island content between the EZH2-bound and EZH2, TRIM37 co-bound promoters.



Extended Data Figure 3 | Control experiments for Fig. 2.

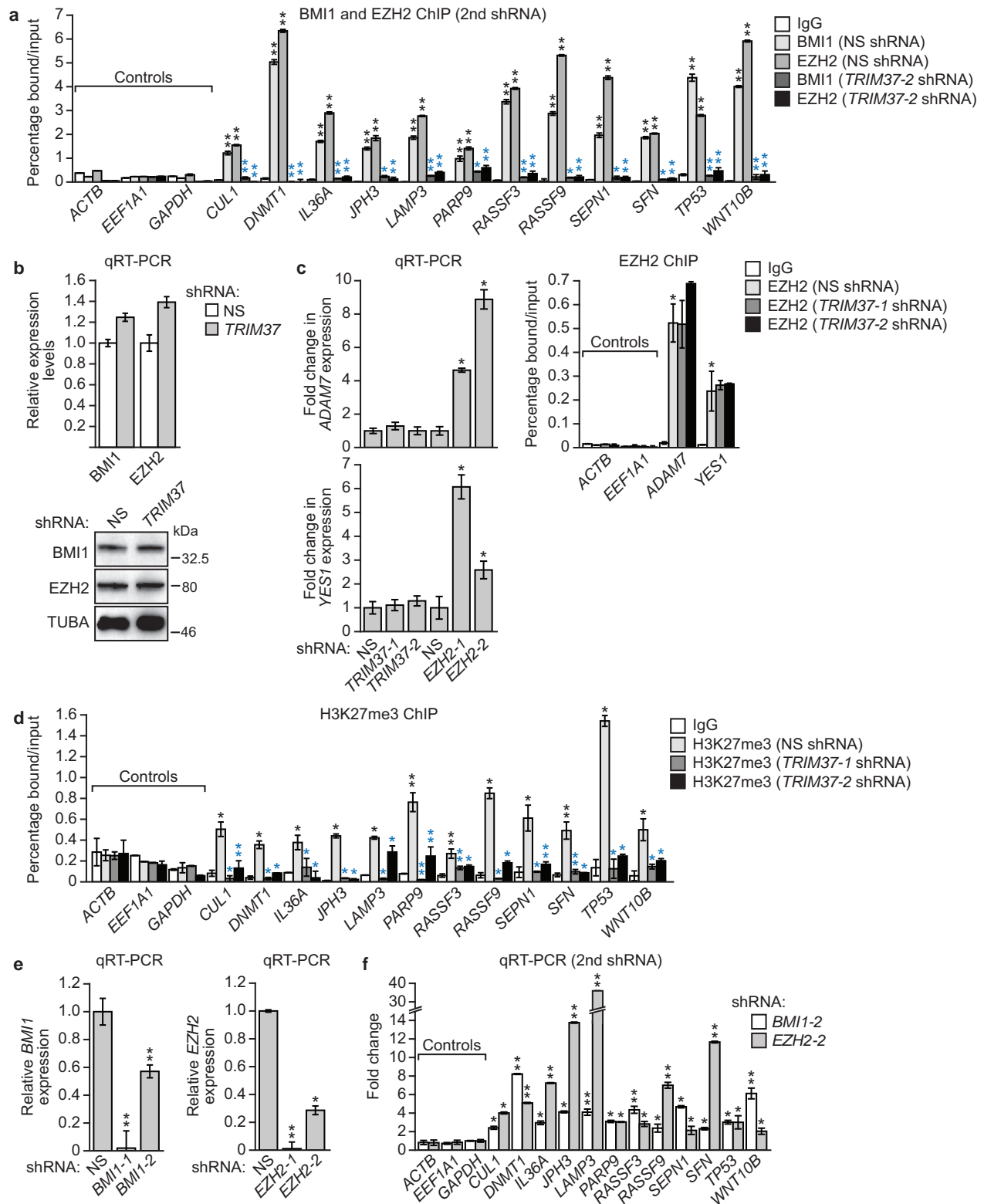
a, ChIP monitoring binding of TRIM37 at promoters of TRIM37 target genes in MCF7 cells expressing an NS or one of two unrelated *TRIM37* shRNAs. As a negative control, TRIM37 binding at three non-TRIM37 target genes, *ACTB*, *EEF1A1* and *GAPDH*, is shown. Black asterisks indicate significance of TRIM37 enrichment compared with the IgG control (from cells expressing an NS shRNA); blue asterisks indicate significant differences in TRIM37 enrichment in cells expressing a *TRIM37* shRNA relative to NS shRNA. Error bars indicate s.d.; $n = 3$ technical replicates of a representative experiment (out of three experiments). The results confirm that TRIM37 occupancy is reduced upon TRIM37 knockdown, demonstrating the specificity of the TRIM37 antibody in ChIP experiments. **b**, ChIP monitoring enrichment of H2Aub at promoters of TRIM37 target genes in MCF7 cells expressing a second *TRIM37* shRNA unrelated to that used in Fig. 2d. The IgG control and H2Aub signal in cells expressing an NS shRNA are the same as those shown in Fig. 2d. Error bars indicate s.d.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **c**, ChIP monitoring enrichment of H2Aub at the promoters of TRIM37 target genes in MCF7 cells expressing an NS or one of two unrelated *RNF2* shRNAs. Error bars indicate s.d.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **d**, qRT-PCR analysis

monitoring expression of TRIM37 target genes in MCF7 cells expressing a second *TRIM37* shRNA unrelated to that used in Fig. 2e. Expression of each gene was normalized to that obtained with an NS shRNA, which was set to 1. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **e**, qRT-PCR analysis monitoring expression of TRIM37 target genes after *RNF2* knockdown in MCF7 cells. Expression of each gene was normalized to that obtained with an NS shRNA, which was set to 1 (indicated by the dotted red line). Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **f**, qRT-PCR analysis monitoring expression of TRIM37 target genes after TRIM37 knockdown in BT474 cells. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **g**, qRT-PCR analysis monitoring expression of *p14ARF* after knockdown of TRIM37 or *EZH2* in BT474 cells. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). The results indicate that knockdown of *EZH2* but not TRIM37 de-represses *p14ARF* expression in BT474 cells. **h**, qRT-PCR analysis monitoring knockdown efficiencies of *TRIM37* (left) and *EZH2* (right) shRNAs in BT474 cells. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). * $P < 0.05$, ** $P < 0.01$.



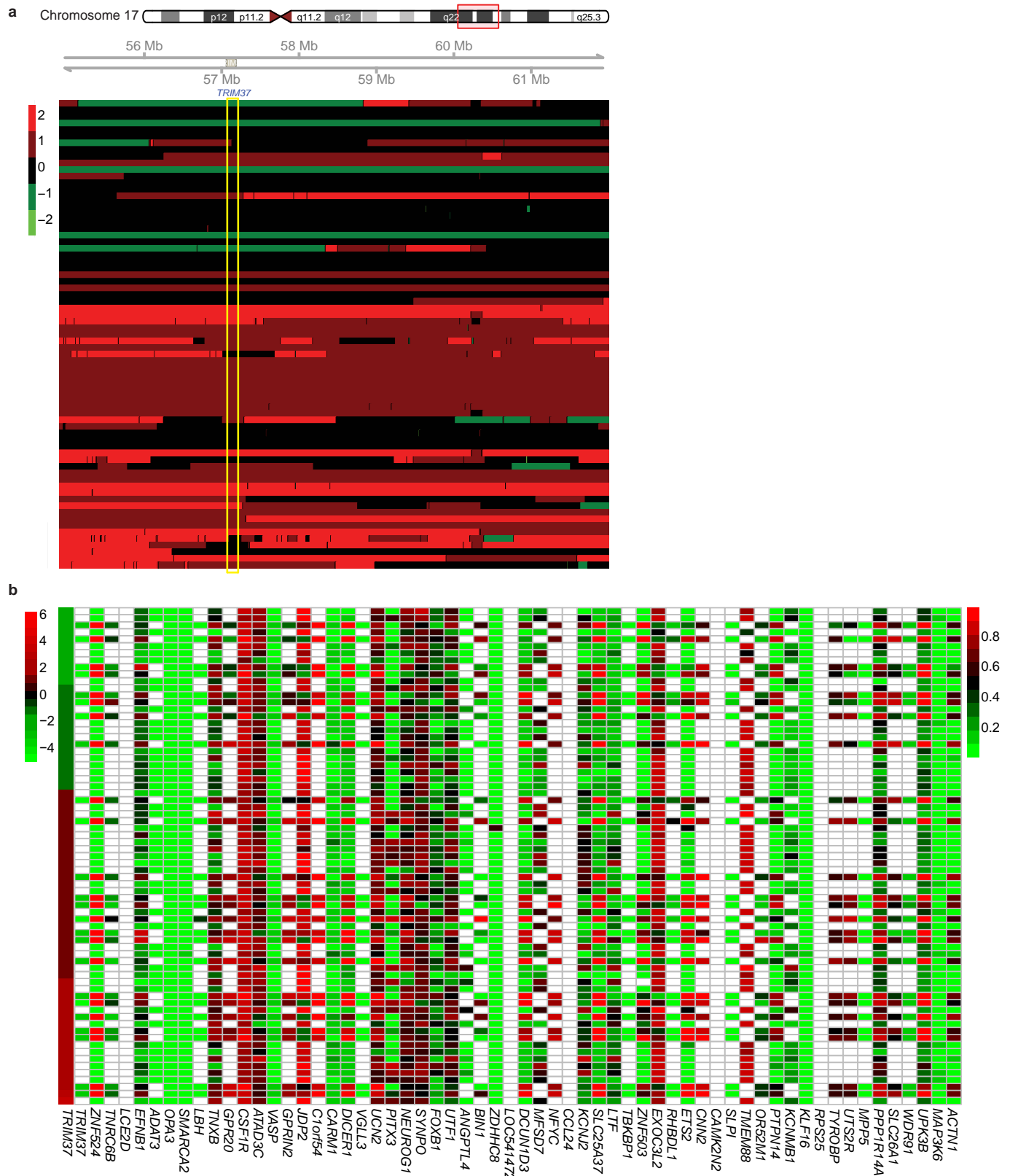
Extended Data Figure 4 | Additional experiments showing interaction between TRIM37 and PRC2 subunits. **a**, Co-immunoprecipitation analysis. MCF7 nuclear extracts were immunoprecipitated with a TRIM37, EZH2 or SUZ12 antibody, or an IgG control, and the immunoprecipitates were analysed for TRIM37, EZH2 or SUZ12 by immunoblotting. **b**, Co-immunoprecipitation analysis. MCF7 nuclear extracts were immunoprecipitated with a TRIM37 or EZH2 antibody, or an IgG control, in the presence of ethidium bromide, and the immunoprecipitates were analysed for TRIM37 or EZH2 by immunoblotting. The results show that interaction between TRIM37 and EZH2 occurs in the presence of ethidium bromide and is thus not mediated by DNA. **c**, Co-immunoprecipitation analysis. Nuclear extracts from MCF7 cells ectopically expressing Flag-TRIM37 were immunoprecipitated with anti-Flag magnetic beads, a EZH2 or SUZ12 antibody, or an IgG control, and the immunoprecipitates were analysed for TRIM37, EZH2 or SUZ12 by immunoblotting. **d**, Mass spectroscopy. Selected results from the liquid chromatography tandem mass spectroscopy analysis listing TRIM37 or PRC2 subunits, and their total spectra score in samples immunoprecipitated using either IgG or a TRIM37 antibody. The probability of interaction, derived from the Scaffold Viewer software, indicates a $\geq 95\%$ probability of interaction in all cases. See also Supplementary Table 2. **e**, Tandem mass spectra of representative peptides of proteins identified in the TRIM37 immunoprecipitate are shown with corresponding spectral counts for each protein shown in the table. **f**, *In vitro* interaction pull-down assay. Purified glutathione-S-transferase (GST)-TRIM37 was incubated with an *in vitro* translated biotinylated PRC2 subunit (indicated on left). GST-TRIM37 was purified using GST-agarose beads, and the presence of the PRC2 subunit

was analysed by immunoblotting with an anti-biotin antibody. The results indicate that TRIM37 interacts strongly with EZH2 and weakly with RBBP4, and does not detectably interact with AEBP2, EED or SUZ12. **g**, GAL4-TRIM37 fusion experiment. Top, schematic diagram of TRIM37 showing the location of the RING, BBOX, coiled coil (CC), MATH and nuclear localization sequence (NLS) motifs. Locations of the motifs were obtained from UniProtKB (http://www.uniprot.org/uniprot/O94972#section_attribute). The Δ BBOX deletion comprises the BBOX and first CC motif (Δ aa 89–235). Middle, left, ChIP monitoring binding of GAL4-TRIM37, EZH2 or SUZ12 to the adenovirus *E1B* promoter containing (GAL4(UAS)-E1B) or lacking (E1B) GAL4-binding sites or to an irrelevant, negative control (NC) DNA region. The indicated GAL4-TRIM37 fusion protein was co-expressed with a plasmid containing or lacking five GAL4-binding sites upstream of the adenovirus *E1B* gene in 293T cells followed by ChIP analysis. Error bars indicate s.d.; $n = 3$ technical replicates of a representative experiment (out of three experiments). The results indicate that the GAL4-TRIM37 fusion protein was able to recruit both EZH2 and SUZ12 to the GAL4-binding sites. Middle, right, ChIP monitoring binding of EZH2 and SUZ12 in the presence of GAL4-TRIM37 deletion mutants. The GAL4-TRIM37 wild-type (WT) samples are the same as those shown in the left panel. The results indicate that the ability of GAL4-TRIM37 to recruit EZH2 and SUZ12 requires the TRIM37 RING domain and NLS (presumably for nuclear entry), but not the BBCC or MATH domains. Bottom, immunoblot analysis monitoring expression of GAL4-TRIM37 fusion proteins using a GAL4 antibody. α -Tubulin (TUBA) was monitored as a loading control.



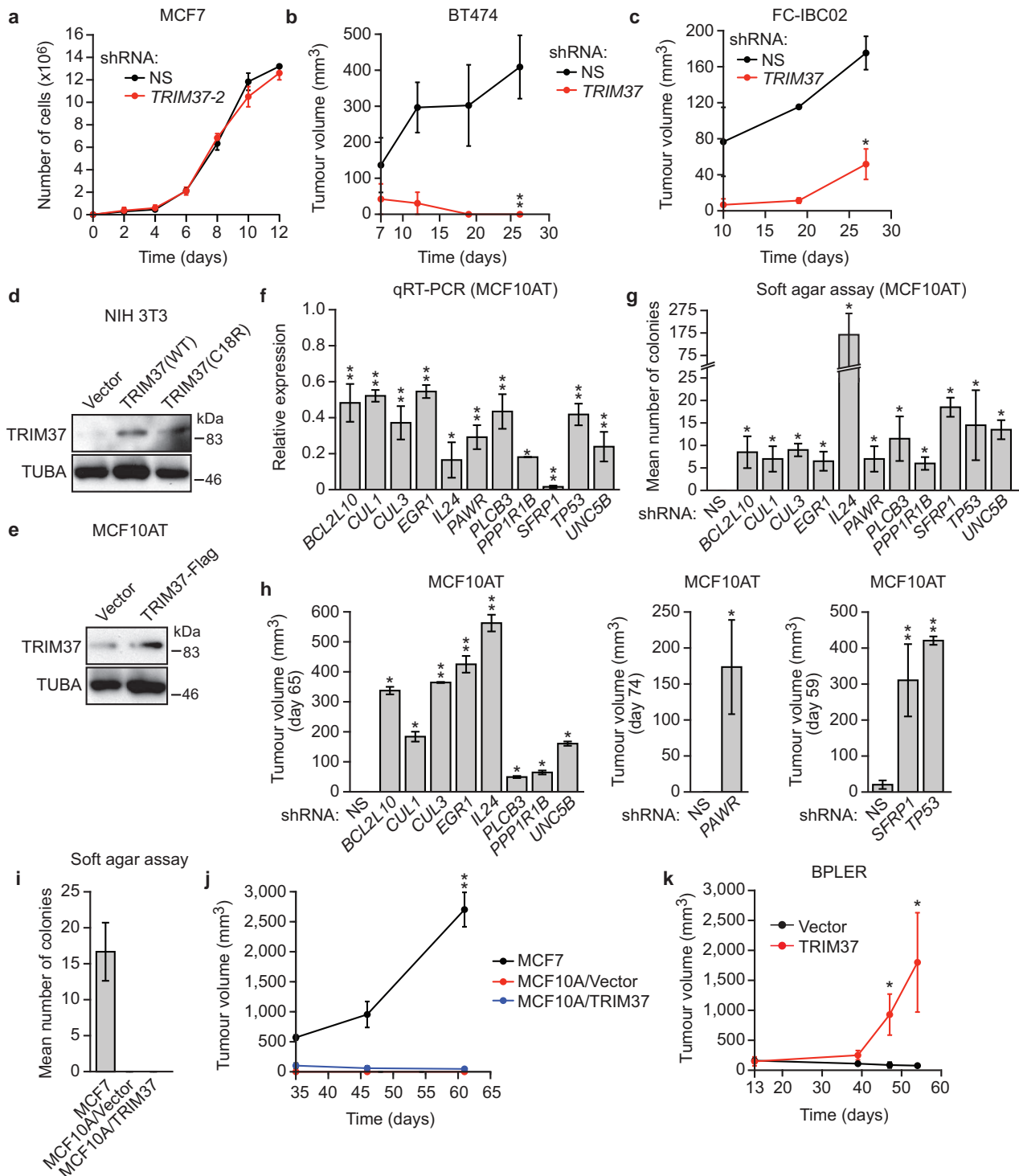
Extended Data Figure 5 | Confirmation of the results of Fig. 3f, h using second, unrelated shRNAs. **a**, ChIP monitoring binding of BMI1 and EZH2 to the promoters of TRIM37 target genes in MCF7 cells expressing an NS shRNA or a second *TRIM37* shRNA unrelated to that used in Fig. 3f. The IgG control and BMI1 and EZH2 signal in cells expressing an NS shRNA are the same as those shown in Fig. 3f. Black asterisks indicate significance of BMI1 or EZH2 enrichment compared with the IgG control (from cells expressing an NS shRNA); blue asterisks indicate significant differences in BMI1 or EZH2 enrichment in cells expressing a *TRIM37* shRNA relative to NS shRNA. Error bars indicate s.d.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **b**, qRT-PCR (top) and immunoblot (bottom) monitoring BMI1 and EZH2 levels in MCF7 cells expressing an NS or *TRIM37* shRNA. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). The results indicate that BMI1 and EZH2 levels are unaffected by *TRIM37* knockdown in MCF7 cells. **c**, Left, qRT-PCR analysis monitoring expression of two genes that are bound by EZH2 but not *TRIM37*, *ADAM7* (top) and *YES1* (bottom), after knockdown of *TRIM37* or EZH2 in MCF7 cells. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). The results indicate that knockdown of EZH2 but not *TRIM37*

de-represses *ADAM7* and *YES1* expression. Right, ChIP monitoring EZH2 enrichment at the promoters of *ADAM7* and *YES1* in MCF7 cells expressing an NS or *TRIM37* shRNA. Error bars indicate s.d.; $n = 3$ technical replicates of a representative experiment (out of three experiments). The results show that knockdown of *TRIM37* has no effect on EZH2 binding at *ADAM7* and *YES1* promoters. Thus, loss of EZH2 binding and de-repression after *TRIM37* knockdown is not general to PRC2-bound genes but rather is selective for *TRIM37* target genes. **d**, ChIP monitoring H3K27me3 enrichment at promoters of *TRIM37* target genes in MCF7 cells expressing an NS or *TRIM37* shRNA. Error bars indicate s.d.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **e**, qRT-PCR analysis monitoring BMI1 (left) and EZH2 (right) knockdown efficiency in MCF7 cells after shRNA-mediated knockdown using two unrelated shRNAs against each gene. Expression of each gene was normalized to that obtained with an NS shRNA, which was set to 1. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **f**, qRT-PCR monitoring *TRIM37* target gene expression in MCF7 cells after knockdown of *BMI1* or *EZH2* using a second shRNA unrelated to that used in Fig. 3h. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). * $P < 0.05$, ** $P < 0.01$.



Extended Data Figure 6 | TRIM37/17q23 copy number and promoter methylation analysis for TRIM37 target genes in human breast cancer samples. **a**, Analysis of 17q23 copy number in a panel of 71 human breast cancer samples; the region corresponding to the *TRIM37* gene is highlighted in yellow. The breast cancer samples shown here are the same as those shown in Fig. 4a. Red indicates increased copy number and green indicates decreased copy number (see Methods). **b**, Promoter methylation analysis. Each of the 60

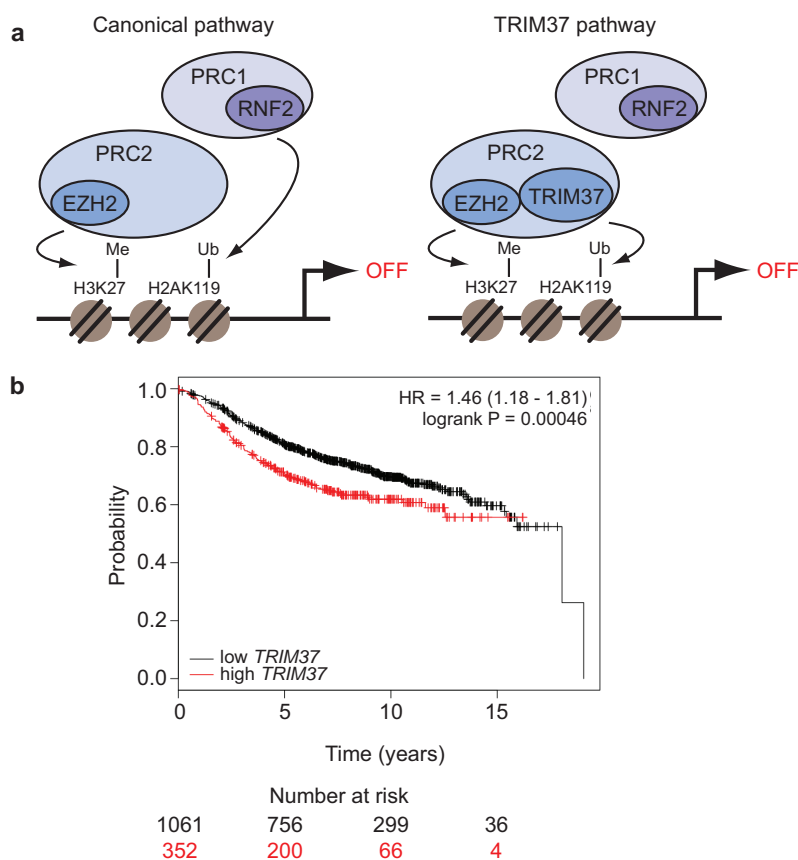
TRIM37 target genes shown in Fig. 4a were analysed for promoter CpG methylation in human breast cancer samples. The level of DNA methylation is shown ranging from low (green) to high (red); white indicates data were not available. The first column is a heat map showing expression of *TRIM37*. The results indicate no significant correlation between *TRIM37* expression and promoter methylation of *TRIM37* target genes.



Extended Data Figure 7 | Control experiments related to Fig. 4.

a, Proliferation of cultured MCF7 cells expressing an NS shRNA or a second *TRIM37* shRNA unrelated to that used in Fig. 4d. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **b**, **c**, Tumour formation in mice subcutaneously injected with BT474 (**b**) or FC-IBC02 (**c**) cells expressing an NS or *TRIM37* shRNA. Error bars indicate s.e.m.; $n = 3$ mice per group. **d**, Immunoblot monitoring *TRIM37* levels in NIH 3T3 cells expressing vector, wild-type (WT) *TRIM37* or *TRIM37*(C18R). α -Tubulin (TUBA) was monitored as a loading control. **e**, Immunoblot monitoring *TRIM37* levels in MCF10AT cells expressing vector or Flag-tagged *TRIM37*. **f**, qRT-PCR analysis monitoring knockdown efficiency of a tumour suppression gene in MCF10AT cells expressing the indicated shRNA. The results are given relative to expression after treatment with an NS shRNA, which was set to 1. Error bars indicate s.d.; $n = 3$ technical replicates of a

representative experiment (out of three experiments). **g**, Soft agar assay monitoring colony formation of MCF10AT cells expressing an NS or the indicated shRNA. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **h**, Tumour formation in mice subcutaneously injected with MCF10AT cells expressing an NS or the indicated shRNA. Error bars indicate s.e.m.; $n = 3$ mice per group. **i**, Soft agar assay monitoring colony formation of MCF10A cells expressing vector or *TRIM37*. Error bars indicate s.e.m.; $n = 3$ technical replicates of a representative experiment (out of three experiments). **j**, Tumour formation in mice subcutaneously injected with MCF10A cells expressing vector or *TRIM37*. Error bars indicate s.e.m.; $n = 3$ mice per group. **k**, Tumour formation in mice subcutaneously injected with BPLER cells expressing vector or *TRIM37*. Error bars indicate s.e.m.; $n = 3$ mice per group. * $P < 0.05$, ** $P < 0.01$.



Extended Data Figure 8 | Schematic model of TRIM37-directed transcriptional repression and Kaplan–Meier analysis comparing survival of breast cancer patients with high or low *TRIM37* expression. **a**, Schematic models for canonical (left) and TRIM37 (right) pathways for target gene silencing. **b**, Kaplan–Meier analysis of survival of patients with low (black) or high (red) *TRIM37* expression. The number of surviving patients at

0-, 5-, 10- and 15-year time points is indicated below the graph. High expression of TRIM37 was significantly correlated with lower survival rate (P value = 0.00046, hazard ratio (HR) = 1.46 with 95% confidence interval 1.18–1.81). The analysis was performed using an online survival analysis tool to analyse the effect of gene expression on breast cancer prognosis using microarray data from 1,809 patients⁴¹.

Extended Data Table 1 | List of catalogue numbers for shRNAs obtained from Open Biosystems/GE Dharmacon or The RNAi Consortium/ Broad Institute

Gene	Catalog number for 1st shRNA	Catalog number for 2nd shRNA
<i>BCL2L10</i>	TRCN0000033595	TRCN0000033596
<i>BMI1</i>	RHS3979-201751271	RHS3979-201751275
<i>CUL1</i>	TRCN0000003391	TRCN0000003392
<i>CUL3</i>	TRCN0000073343	TRCN0000073344
<i>EGR1</i>	TRCN0000013073	TRCN0000013074
<i>EZH2</i>	TRCN0000040074,	TRCN0000040076,
<i>IL24</i>	RHS3979-9607463	RHS3979-9607465
<i>PAWR</i>	TRCN0000058443	TRCN0000058444
<i>PLCB3</i>	TRCN0000020304	TRCN0000020305
<i>PPP1R1B</i>	TRCN0000000434	TRCN0000000433
<i>Rnf2</i> (mouse)	TRCN0000052879	TRCN0000052880
<i>RNF2</i>	TRCN0000040579	TRCN0000040580
<i>SFRP1</i>	RMM3981-201769168	RHS3979-9601103
<i>TP53</i>	TRCN0000062172	TRCN0000062172
<i>Trim37</i> (mouse)*	TRCN0000018866	TRCN0000018869
	TGCTGTTGACAGTGAGCGACCGGGATCAGCACTG- TATATTAGTGAAGCCACAGATGTAATA- TACCAGTGCTGATCCCGGCTGCCTACTGCCTCGGA	TGCTGTTGACAGTGAGCGAGGGATCAGCACTGG- TATATTAGTGAAGCCACAGATGTATA- ATATACCAGTGCTGATCCCGTGCCTACTGCCTCGGA
<i>TRIM37</i> (human)	RHS4430-200178645	RHS4430-200209374
<i>UNC5B</i>	TRCN0000061816	TRCN0000061813

*These shRNAs are no longer commercially available from Open Biosystems/GE Dharmacon, and thus shRNA sequences are provided in lieu of catalogue numbers.

Extended Data Table 2 | List of primers used for quantitative real-time RT-PCR, ChIP and vector construction

	Gene	Forward primer (5' → 3')	Reverse primer (5' → 3')
qRT-PCR	<i>ACTB</i>	ACGTTGACATCCGTAAGACCT	GCAGTAATCTCCTTCTGCATCC
	<i>ACTN1</i>	TCCATCGGAGCCGAAGAAATC	GTGTCGGTGGATCAAAGCACA
	<i>ADAM7</i>	GGTGCCGTATGGTGCCAATTA	CCCCAAATTCGGTTCTCTAGTT
	<i>ASS1</i>	TCCGTGGTTCTGGCCTACA	GGCTTCTCGAAGTCTTCCTT
	<i>BCL2L10</i>	CGGAAGAAGCGCCATCTCC	AGGCGTCTCTTCACTTCAG
	<i>BMI1</i>	TTCTGCTGATGCTGCCAATG	TCCGATCCAATCTGTTCTGGTC
	<i>CDT1</i>	ATGTCAAGGAGCACCACAAG	ACTTCATCCACGTTGAAGCG
	<i>CKB</i>	ACAACAACACATGGCCAAG	AGCCCCACGGTCATGATGAC
	<i>CUL1</i>	AACAGTGCAAGTGACGATGC	TTGCTCACGCCAATGTCTTG
	<i>CUL3</i>	TAAACCAACACAGCGGGTTC	TCACCTTGTTTGGCAGCAAC
	<i>DDR1</i>	TTCCCCCTTAATGTGCGTAAGG	AGCAGCCGAATGATGTTTGG
	<i>DNMT1</i>	GAGCTACCACGACGACATCA	CGAGGAAGTAGAAGCGGTTG
	<i>EEF1A1</i>	GAAGACCCACATCAACATCGT	CTCCGCATTTGTAGATGAGGTG
	<i>EGR1</i>	CAGCACCTTCAACCCTCAGG	GAGATGTACAGGAAAAGACTCTGC
	<i>EZH2</i>	AGTTCGTGCCCTTGTGTGATAG	TGGTGTGCACTGTGCTTTG
	<i>GAPDH</i>	TGCACCACCACTGCTTAGC	GGCATGGACTGTGGTCATGAG
	<i>GRIN1</i>	ACGCCATCTCTAGTTAGCCATC	GCACGGGTATGCGGTAGAAG
	<i>HEYL</i>	GGCTGCTTACGTGGCTGTT	GACCCAGGAGTGGTAGAGCAT
	<i>IL24</i>	TTCTGGGCTGTGAAGACAC	TTTCTTGCAGACGTTCTGC
	<i>IL36A</i>	CCAGACGCTCATAGCAGTCC	AGATGGGGTTCCTCTGTCTT
	<i>JPH3</i>	AGAATGGGGATGAGCTCAAGTC	TTTTGCTACCCGACATCTC
	<i>LAMP3</i>	GCGTCCCTGGCCGTAATTT	TGCTTGCTTAGCTGGTTGCT
	<i>LTBP3</i>	TCCCCAGGGCTACAAGAGG	AGACACAGCGATAGGAGCCA
	<i>PARD6G</i>	GTCTCAGACCTTGCGATTCTAC	GGTCCAGAGAGAACCCTTCGGA
	<i>PARP9</i>	TGTTTGTGGCCATTTCTGTG	ACCATGTTCAACGACACAAC
	<i>PAWR</i>	TGATGAAGCAGGGCAGAAAAG	AAACTGTTCTAGGTGGCTCCGT
	<i>PLCB3</i>	TTCAATTACCCACGGCTTAC	TCTTTGCCCTGTTGCTTTGCC
	<i>PPP1R1B</i>	CAAGTCGAAGAGACCCAAACCC	GCCTGGTTCTCATTCAAATGTCT
	<i>PROP1</i>	GCCCCAGTGCAGTTGGAACA	TGCTTCCGTTGCTTAGCTCTG
	<i>P14ARF</i>	CCCTCGTGCTGATGCTACTG	ACCTGGTCTTCTAGGAAGCGG
	<i>RASSF3</i>	GCCTGCAAGCTCTCAGACC	TAGGCTGTGTAGCGCCTCTT
	<i>RASSF9</i>	GAGGACCTGAGCGAAAGTGAT	TCTGGATGCCACTCAATAGAGA
	<i>RNF2</i>	GTGCAGACAACCGAACTCA	TTGCCACTTCTAAGGGCTGT
	<i>SEPN1</i>	CCTGACCCTAGCGAGGAGAC	GGCTGTCCAGTTTCGGAGG
	<i>SFN</i>	TGACGACAAGAAGCGCATCAT	GTAGTGGAAGACGGAAAAGTTCA
	<i>SFRP1</i>	TCATGCAGTTCTTCGGCTTC	TTGTCAAGGGAGGACACAC
	<i>TBKBP1</i>	GGCGGCTAGAAGAGGCTTTG	CTCCGTCAGCTCGGTGTAG
	<i>TP53</i>	ACTTGTCCGCTCTTGAAGCTAC	GATGCGGAGAACTTTTGAACA
	<i>TRAF2</i>	TGCCCAAGTTCCCTTAAAC	TTGCCACAAGTCTTGACGTG
	<i>TRIM37</i>	AACAGAGCGTGGAGAGCAT	CTTCTGCCCAACGACAATT
	<i>UNC5B</i>	GTCGGACACTGCCAACTATAC	CCGCCATTACGTAGACGAT
	<i>VCL</i>	AAACTGAACAGGCCAAAGG	TTTGCCTGCACAGAGTTTAC
	<i>WNT10B</i>	GTGAGCGAGACCCCACTATG	CACCTCTGAACCTTGACATCAT
	<i>YES1</i>	GCCTGTCAAGTGTGAG	AAAGGCGTTACCCCTGAGGAT
ChIP	<i>Actb</i> (mouse)	CGAAGCCGGTGAGTGAGC	ACCGCCGAGACCGCGTCC
	<i>ACTB</i>	CCGTTCCGAAAGTTGCCTT	CGCCGCCGGGTTTTATA
	<i>ADAM7</i>	AACTGGGAGCAAGTGAAGT	CTGCTCCAAGGACGACTTTTTT
	<i>BCL2L10</i>	AGGAGGTTGTGGTTCAGACC	ACTTAGCAGCCAGAGTGATCC
	<i>CDT1</i>	GGGATTACAGGCGTGAACCA	ATCCGCTCTCTTGGAACTGC
	<i>CKB</i>	CAGGTGTCCCTTAAGCCTGG	CCCTTCAGGTTTTGTGGGGT
	<i>CUL1</i>	CGAGGGAGGCTGTAAAACCTG	CGTGGGAAGCTTCTACCTG
	<i>DDR1</i>	ACCTGCTTCTTCTCTGTGC	CCCCAAACGAGAACCCTCAA
	<i>DNMT1</i>	CCACGCTTGGCTAATTTTGG	ATCACGAGGTGAGGAGTTCTG
	<i>EEF1A</i>	AGCTCGGATCAAGAATCAC	TGGCCTCTTACGGGTATG
	<i>EGR1</i>	CTGGGAGTGGAGAGGGAAC	AATCCGAAAGAGGGGCTTCC
	<i>Fas</i> (CP/TSS)	GCCGCTGTGCAAGTGGTGA	CTGTGTGTTGGGACGCTCGGG
	<i>Fas</i> (1 kb)	GGCTATAGATCACCTTATGTA	GCAGTTAACTCAGGGACCAAG
	<i>Fas</i> (2 kb)	GCGTTGCCATAGCATGAAC	GAGTTAGGGGACCATAGTCA
	<i>Gapdh</i> (mouse)	CACGCTCGGTGCGTGCCC	GCTCAGCCAGTCCCAGCC
	<i>GAPDH</i>	TCCGGGTGATGCTTTTCTAG	TTTGGCGTGGAAATGTCCTTTT
	<i>IL24</i>	TTCTTCGCTGGGACTACTGTG	AACAACACAGAGGTGCTTGG
	<i>IL36A</i>	TCAACTATGTGTGGCTTGTGC	GCTAAACTCATTCCACAAGAAAGC
	<i>JPH3</i>	GTTTGCAGAAAGGGCTCAGT	AGGAACCTGCGCTATGGGAAG
	<i>LAMP3</i>	AGCACCTGGACAGCAATTAC	AGGTGAGGTTGTGAGGTGAATG
	<i>PARP9</i>	GTGTGGTTCTGGCTAAATGCAC	TGCGGCACATTTGTACAAGG
	<i>PAWR</i>	TGTTGTGGGAAACGTATGGC	TTAGGTAACACGCGCTCAGC
	<i>PLCB3</i>	TGTGAAGGTCTCCAGCCATG	ATGGGAACCTGGACAGGAATCTG
	<i>RASSF3</i>	TAATCCAGCTACTTGGGAAGC	TTTTGCTCTGTCAACCAGAC
	<i>RASSF9</i>	ATCATCGCCACCTGGTATCAC	GCAAAACAGAGGGAACAACATGG
	<i>SEPN1</i>	AGGCGCAGTGTATGTGTAC	TCGGAGGAAGGAAATGACAGAC
	<i>SFN</i>	ACAGGCCACCAACTTGAAAG	TAACAAGCAGCCCTCTGAAGC
	<i>SFRP1</i>	CCCTCCACCTCAAGCCAAAT	ACAAGTCCTCCTGCCATTG
	<i>TP53</i>	TTGCCAGGAAAGATGCACAG	ATGTGGCAGTTGTGTTGCTC
	<i>TRAF2</i>	GCGCGCTCTTTGAGGAAAAA	CAGCCCTCAGGAAGCTGTAG
	<i>WNT10B</i>	TTCTTCTGCTCCACATCTC	AGCACTTCTCGGAAACCTG
	<i>YES1</i>	TGATGAGGGTGTGAGTAACGC	TGGTTGGCTGTTTGTCTCTC
GAL4 recruitment	<i>GAL4-UAS</i>	GACACTATAGAATCGAGCAGCTG	ATGAATTGAGCTCGGTACCC
	<i>TRIM37</i>	AACAGAGCGTGGAGAGCATT	CTTCTGCCCAACGACAATT
Vector construction	<i>TRIM37-C18R</i>	CATACGTATGAAACATCGGAAACCTCAGC	CATACGTATGAGGAAATTCGGGGATG
	<i>H2A-KKRR</i> (N terminus)	ATGTCTGGTCTGGGCAACAA	ATGCGCGTCTCTCTGTTGTC
	<i>H2A-KKRR</i> (C terminus)	(P)-CATCCCTCGTCACTCCAG	TGGTGGCTCTAAAAGAGCC
	Cloning TRIM37 into pGEX4T-1	GAATTTCATGGATGAACAGAGCGTGGA	GCGGCCGCTCTTCCACTATTTT
	Cloning TRIM37 into pCMX-Gal4	GGTACCAAAATGGATGAACAGAGCGTGGA	GCTAGCTCTTCCACTATTTT
	<i>TRIM37ΔRING</i>	GAATTTCATGGATGAACAGAGCGTGGA	CTCCAGCTACGAGAAGTAGTAAA
		TACGAGAACTAGTAAA	
	<i>TRIM37ΔBBOX</i>	(P)-ATGTTTGGTGAGACTGCAGAG	ATTTATGAGCAACACGTCACTAAA
	<i>TRIM37ΔCC</i>	(P)-AGTGACGTTGCTCATAAATTTCT	GAGTTGATATCTAAGAGCTCAGAGA
	<i>TRIM37ΔMATH</i>	(P)-AGTAGCTGAATCGTAAGATGGCA	GCAAAATGAAGGATACTTGAATCCACA
	<i>TRIM37ΔNLS</i>	(P)-CAAGCCACTGAAACTGCAACC	ATGGTCACCTTGGGGGCTAAT

Piezo2 is the major transducer of mechanical forces for touch sensation in mice

Sanjeev S. Ranade¹, Seung-Hyun Woo¹, Adrienne E. Dubin¹, Rabih A. Moshourab^{2,3}, Christiane Wetzel², Matt Petrus⁴, Jayanti Mathur⁴, Valérie Bégay², Bertrand Coste^{1†}, James Mainquist⁴, A. J. Wilson⁴, Allain G. Francisco¹, Kritika Reddy¹, Zhaozhu Qiu^{1,4}, John N. Wood⁵, Gary R. Lewin² & Ardem Patapoutian¹

The sense of touch provides critical information about our physical environment by transforming mechanical energy into electrical signals¹. It is postulated that mechanically activated cation channels initiate touch sensation, but the identity of these molecules in mammals has been elusive². Piezo2 is a rapidly adapting, mechanically activated ion channel expressed in a subset of sensory neurons of the dorsal root ganglion and in cutaneous mechanoreceptors known as Merkel-cell–neurite complexes^{3,4}. It has been demonstrated that Merkel cells have a role in vertebrate mechanosensation using Piezo2, particularly in shaping the type of current sent by the innervating sensory neuron^{4–6}; however, major aspects of touch sensation remain intact without Merkel cell activity^{4,7}. Here we show that mice lacking Piezo2 in both adult sensory neurons and Merkel cells exhibit a profound loss of touch sensation. We precisely localize Piezo2 to the peripheral endings of a broad range of low-threshold mechanoreceptors that innervate both hairy and glabrous skin. Most rapidly adapting, mechanically activated currents in dorsal root ganglion neuronal cultures are absent in *Piezo2* conditional knockout mice, and *ex vivo* skin nerve preparation studies show that the mechanosensitivity of low-threshold mechanoreceptors strongly depends on Piezo2. This cellular phenotype correlates with an unprecedented behavioural phenotype: an almost complete deficit in light-touch sensation in multiple behavioural assays, without affecting other somatosensory functions. Our results highlight that a single ion channel that displays rapidly adapting, mechanically activated currents *in vitro* is responsible for the mechanosensitivity of most low-threshold mechanoreceptor subtypes involved in innocuous touch sensation. Notably, we find that touch and pain sensation are separable, suggesting that as-yet-unknown mechanically activated ion channel(s) must account for noxious (painful) mechanosensation.

Dorsal root ganglion (DRG) neurons have pseudounipolar axons that terminate in the skin where they form specialized mechanoreceptors that are tuned to detect mechanical forces such as stretch, indentation and vibration¹. A diverse set of low-threshold mechanoreceptors (LTMRs) are distributed within hairy and glabrous skin. Lanceolate and circumferential endings that contain a mixed population of A β , A δ and C LTMRs are specific to hairy skin, whereas various corpuscles that consist of rapidly adapting A β LTMRs are unique to glabrous skin^{1,8}. Merkel-cell–neurite complexes that mediate slowly adapting responses in A β fibres are present in both skin types⁴. Together, these structurally diverse LTMRs detect mechanical stimuli relevant to innocuous touch sensation.

We had previously shown that Piezo2 was expressed in Merkel-cell–neurite complexes and that Merkel cells partly contributed to slowly adapting type I firing^{4,5,9}. Whether Piezo2 was also expressed in other cutaneous mechanoreceptors and whether it functioned as the primary mechanotransduction ion channel was not known. We therefore used a recently generated mouse line, *Piezo2-GFP*, where green fluorescent

protein (GFP) was fused to the carboxy terminus of *Piezo2*, to detect localization of Piezo2 at the site of mechanical transduction: in the nerve terminals of sensory neurons that project to the skin⁴. In hairy skin, GFP expression was found in both lanceolate endings and circumferential fibres that have varying degrees of myelination and wrap around the bulge region of hair follicles (Fig. 1a, b). Consistent with our previous analysis, Piezo2 was expressed in Merkel cells in both hairy and glabrous skin (Fig. 1c)^{4,5}. In glabrous skin, Piezo2 was also expressed in Meissner's corpuscles (Fig. 1d). Overall, Piezo2 expression is found in a broad range of LTMRs that sense mechanical stimuli relevant to touch sensation.

Constitutive deletion of Piezo2 led to perinatal lethality; therefore, we used an inducible strategy to delete Piezo2 in adult mice¹⁰. The *Advil-creERT2* mouse line allowed for tamoxifen-induced activation of Cre recombinase under the *advillin* promoter in sensory neurons and in epidermal Merkel cells¹¹. We first characterized the expression overlap in DRGs between *Advil-creERT2* and Piezo2 by mating the *Advil-creERT2* mouse to the Ai9 tdTomato reporter line. In agreement with a previous report, we found that 87% of DRG neurons express tdTomato (766 of 876 total cells) (Fig. 2a). Co-expression analysis of tdTomato epifluorescence with anti-Piezo2 antibody staining showed that 82% of Piezo2⁺ cells were also tdTomato⁺ (343 of 419 cells out of 876 total) (Fig. 2a)⁴. We also detected expression of tdTomato in epidermal Merkel cells (data not shown), indicating the *Advil-creERT2* mouse line would lead to deletion of Piezo2 in all cell types either proposed or known to be relevant to somatosensory transduction¹¹. These data further indicated that while *Advil-creERT2* would ablate Piezo2 in most DRG neurons, a small number of Piezo2⁺/Cre[−] neurons might still remain intact. We then mated the *Advil-creERT2* mouse to a previously generated conditional knockout (CKO) mouse line, *Piezo2^{fl}*, and found that Piezo2 expression levels were similar between genotypes before tamoxifen injections (Fig. 2b, top). After tamoxifen injections, Piezo2 immunostaining in *Piezo2^{CKO}* mice showed a marked decrease in the number of Piezo2-positive DRG neurons compared to wild-type (*Piezo2^{WT}*) mice (Fig. 2b, bottom) and quantitative real-time PCR (qPCR) analysis revealed a ~90% deletion of overall *Piezo2* transcripts in isolated DRG neurons (Fig. 2c).

Previous reports had shown that cultured DRG neurons that were transfected with small interfering RNA (siRNA) for *Piezo2* showed a selective decrease in rapidly adapting, mechanically activated currents^{3,12}. We determined the sensitivity of cultured DRG neurons from *Piezo2^{CKO}* mice to mechanical indentation using a piezoelectrically actuated blunt glass probe^{10,13,14}. DRGs from *Piezo2^{CKO}* mice had dramatically fewer neurons with rapidly adapting whole-cell current responses compared to controls (Fig. 2d and Extended Data Fig. 1a), and a corresponding increase in the proportion of mechanically insensitive neurons. There were no significant effects on intermediately adapting and slowly adapting currents (Fig. 2d and Extended Data Fig. 1b, c), and there were no differences between genotypes for access resistance, membrane resistance

¹Howard Hughes Medical Institute, Molecular and Cellular Neuroscience, Dorris Neuroscience Center, The Scripps Research Institute, La Jolla, California 92037, USA. ²Department of Neuroscience, Max-Delbrück Center for Molecular Medicine, Robert-Rössle Straße 10, D-13092 Berlin, Germany. ³Klinik für Anästhesiologie mit Schwerpunkt Operative Intensivmedizin, Campus Charité Mitte and Virchow-Klinikum Charité, Universitätsmedizin Berlin, Augustenburgerplatz 1, 13353 Berlin, Germany. ⁴Genomics Institute of the Novartis Research Foundation, San Diego, California 92121, USA. ⁵Molecular Nociception Group, Wolfson Institute for Biomedical Research, University College London, London WC1E 6BT, UK. [†]Present address: Ion Channels and Sensory Transduction group, Aix Marseille University, CNRS, CRN2M-UMR 7286, 13344 Marseille, France.

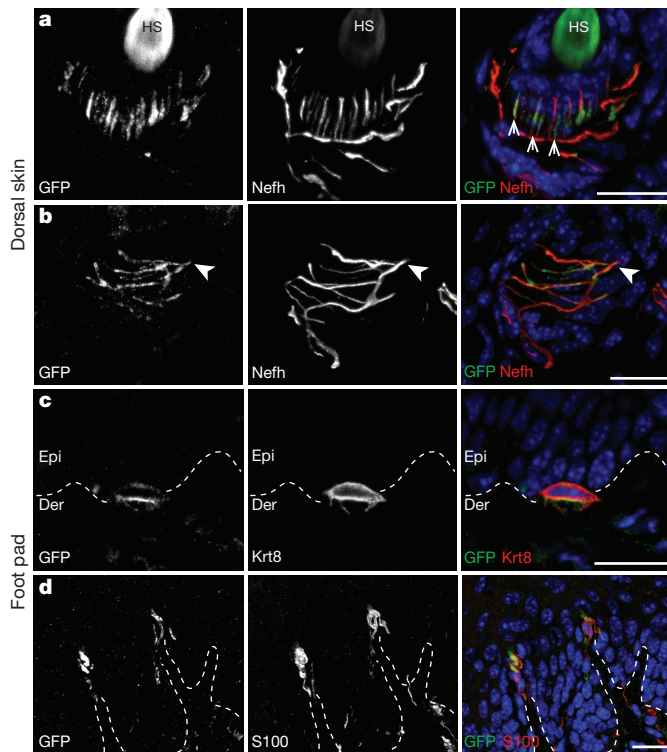


Figure 1 | Piezo2 is localized at the nerve terminals of sensory neurons that innervate the skin. **a, b,** Representative images of immunostaining of GFP, Neff and DAPI (4',6-diamidino-2-phenylindole, blue) in the hair follicle of dorsal skin of *Piezo2*-GFP mice. GFP staining indicates localization of the Piezo2-GFP fusion protein and Neff marks myelinated neurons. **c,** Representative image of immunostaining of GFP, Krt8 (a specific marker of Merkel cells) and DAPI (blue) in glabrous skin of *Piezo2*-GFP mice. **d,** Representative images of immunostaining of GFP and S100 (a marker of Schwann cells) in glabrous skin. Arrows mark lanceolate endings (**a**) and arrowheads mark circumferential fibres (**b**). Dashed lines demarcate the epidermal-dermal junction (**c** and **d**). Der, dermis; Epi, epidermis; HS, hair shaft. Scale bars, 20 μ m.

and resting membrane potential (Extended Data Table 1). Notably, we observed a wide range of low and high apparent thresholds for rapidly adapting currents in *Piezo2*^{WT} neurons, while none of the remaining rapidly adapting responses in *Piezo2*^{CKO} neurons had low apparent thresholds (Extended Data Fig. 2). Whether the rapidly adapting activity in sensory neurons of *Piezo2*^{CKO} mice is mediated by any remaining Piezo2 or another mechanically activated channel is not known. Nonetheless, these results suggest that Piezo2 accounts for the majority of rapidly adapting currents in cultured DRG neurons.

In order to confirm that loss of Piezo2 did not affect the integrity of DRG neurons, we performed immunostaining analyses of DRG neurons using markers of various subpopulations. The expression patterns of neurofilament heavy polypeptide (Neff), thymidine hydroxylase (TH) and calcitonin gene-related peptide (CGRP) were unaffected in *Piezo2*^{CKO} DRGs relative to *Piezo2*^{WT} DRGs (Extended Data Fig. 3a–f). Moreover, *Piezo2*^{WT} mice, as well as *Piezo2*^{CKO} mice, showed normal cutaneous mechanoreceptor formation including Merkel-cell-neurite complexes (Extended Data Fig. 4a, d), lanceolate and circumferential endings (Extended Data Fig. 4b, e, f), and Meissner's corpuscles (Extended Data Fig. 4c, g). We also performed calcium imaging in cultured DRG neurons and found no significant differences in the number of capsaicin-responsive cells between the two genotypes (data not shown).

Loss of mechanically activated currents in cultured DRG neurons is expected to be associated with a loss of mechanosensitivity of sensory fibres in the skin¹⁵. We tested this using an *ex vivo* skin nerve preparation in which mechanically insensitive fibres were identified with an

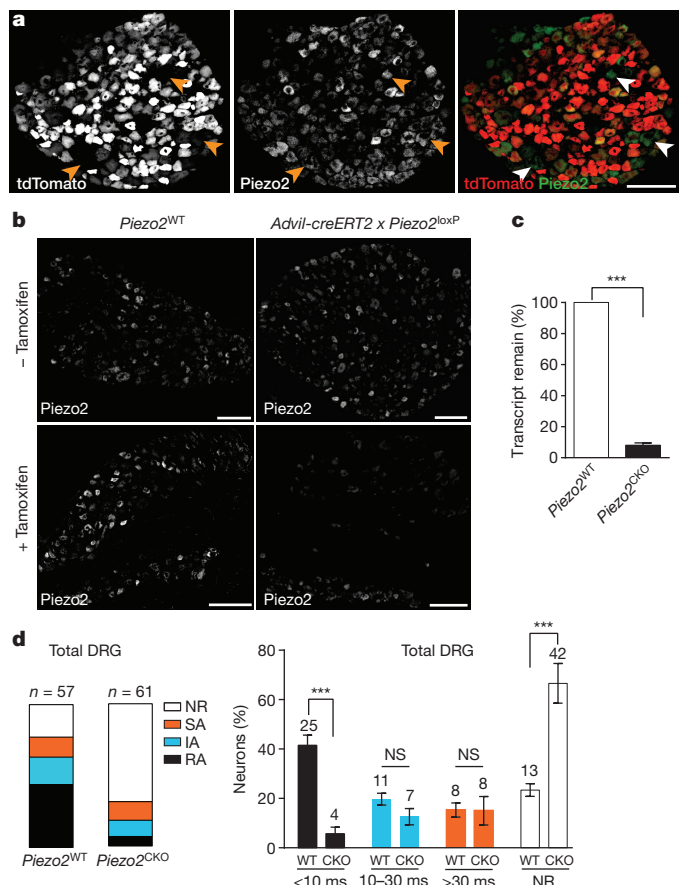


Figure 2 | Advil-creERT2 mediates efficient deletion of Piezo2 and leads to specific loss of rapidly adapting, mechanically activated currents in cultured DRG neurons. **a,** Immunostaining of DRG neurons from tamoxifen-treated *Advil-creERT2* × *Ai9* mice for tdTomato epifluorescence and anti-Piezo2 antibody. Arrowheads indicate the relatively few tdTomato⁺/*Piezo2*⁺ neurons that would presumably not be deleted in *Piezo2*^{CKO} mice. **b,** Representative images of immunostaining using anti-Piezo2 antibody in DRGs from *Advil-creERT2* × *Piezo2*^{loxP} mice before and after tamoxifen treatment. **c,** qPCR of *Piezo2* from isolated DRGs from *Piezo2*^{WT} and *Piezo2*^{CKO} mice with *Piezo2*^{WT} values normalized at 100% ($n = 3$ independent experiments, $P < 0.0001$, Student's *t*-test). **d,** The proportion of DRG neurons responding with rapidly (RA, $\tau_{\text{inact}} < 10$ ms), intermediately (IA, $\tau_{\text{inact}} 10$ –30 ms) and slowly (SA, $\tau_{\text{inact}} > 30$ ms) adapting mechanically activated currents from *Piezo2*^{WT} and *Piezo2*^{CKO} littermates. NR, non-responsive to displacements of at least one-third cell diameter; NS, not significant. Results from $n = 3$ independent experiments. Error bars represent s.e.m., *** $P < 0.0001$ for rapidly adapting and *** $P < 0.0005$ for non-responsive, Student's *t*-test. Scale bars, 100 μ m (**a, b**).

electrical search technique (Extended Data Table 2)^{15,16}. In *Piezo2*^{CKO} mice, 50% of the A β fibres had no detectable mechanosensitivity compared to less than 10% of A β fibres from *Piezo2*^{WT} mice (Fig. 3a). There was a non-significant decrease in the proportion of A δ fibres in *Piezo2*^{CKO} lacking mechanosensitivity, and there was no significant loss of mechanosensitivity in C fibres (<10% in both genotypes) (Fig. 3a). Despite the marked loss of mechanosensitivity, the axonal conduction velocities of A β , A δ and C fibres were unchanged in *Piezo2*^{CKO} mice compared to controls (Extended Data Fig. 5a).

We also investigated A β fibres with a mechanosensitive receptive field using a series of ramp and hold force stimuli with ramp phases of different velocities (Fig. 3b, c). Slowly adapting mechanoreceptors associated with Merkel cells are known to have a characteristic irregular firing during static displacement and a prominent dynamic discharge during the ramp¹⁷. We showed previously that deletion of Piezo2 in Merkel cells led to reduced firing in the static phase but not the dynamic phase⁴.

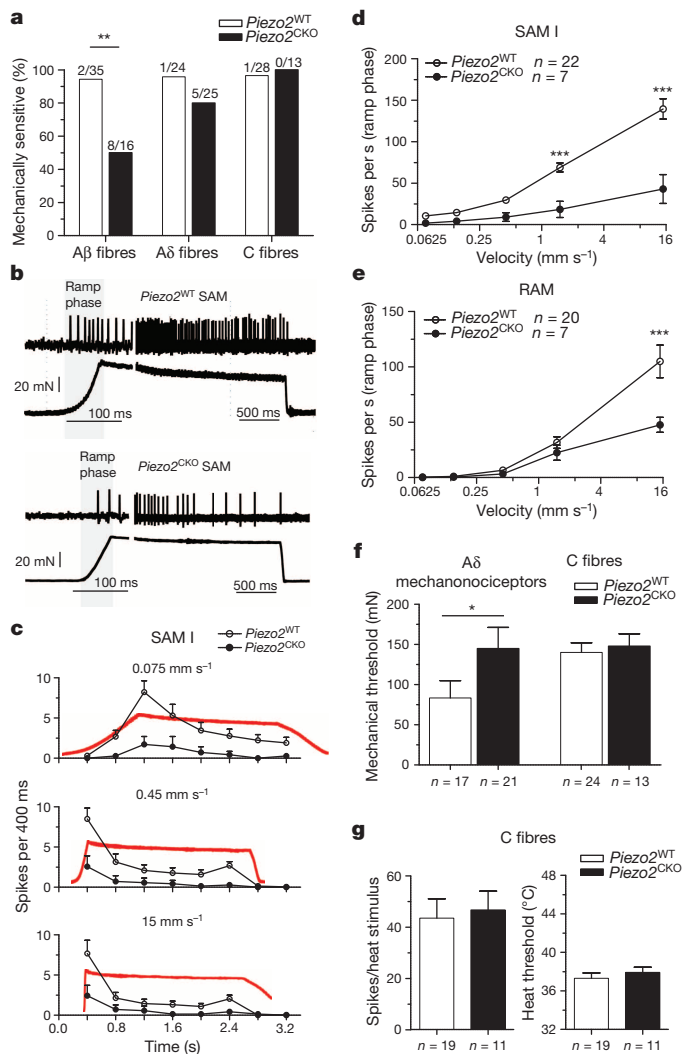


Figure 3 | Piezo2 is required for mechanoreceptor function in *ex vivo* skin nerve preparation. **a**, Proportions of mechanically sensitive Aβ, Aδ and C fibres in *Piezo2*^{CKO} mice compared to *Piezo2*^{WT} controls, one-sided Fisher's exact test, ***P* < 0.01. **b**, Typical examples of SAM I mechanoreceptor responses from *Piezo2*^{CKO} and *Piezo2*^{WT} mice. **c**, Mean discharge rates (400-ms bins) during the course of ramp and hold stimuli with different onset velocities for SAM I receptors. Note the almost complete lack of dynamic ramp discharge in SAM I receptors recorded from *Piezo2*^{CKO} mice. **d**, Normally, SAM I discharge rates increase with increasing ramp speed, however SAM I receptors from *Piezo2*^{CKO} mice showed a strong reduction in velocity sensitivity with firing discharge increasing only marginally with increasing stimulus velocity (repeated measures ANOVA, *F* = 19.69, *P* < 0.0001, with Bonferroni post hoc test, ****P* < 0.001). **e**, Rapidly adapting afferents also displayed reduced velocity sensitivity in *Piezo2*^{CKO} mutant mice (repeated measures ANOVA, *F* = 4.36, *P* < 0.05, with Bonferroni post hoc test, ****P* < 0.001). **f**, Mean force thresholds for nociceptor discharge. Individual nociceptors threshold were calculated by averaging the thresholds of first spikes occurring during the stimulus ramp phase for all six stimuli applied. Note significant elevation of mechanical thresholds for Aδ mechanonociceptors in *Piezo2*^{CKO} mice (Mann–Whitney test, **P* < 0.05), but no change in C-fibre mechanical thresholds. **g**, Discharge rates and heat thresholds to standard noxious heat ramps did not differ between polymodal C fibres recorded from *Piezo2*^{CKO} and *Piezo2*^{WT} mice. Data are presented as mean ± s.e.m.

Here, most slowly adapting mechanoreceptors encountered in *Piezo2*^{WT} (22 of 31 fibres) and *Piezo2*^{CKO} (7 of 10 fibres) could be classified as probably slowly adapting type I fibres (SAM I) associated with Merkel cells (Extended Data Table 2)^{17,18}. Notably, the characteristic dynamic response of SAM I fibres recorded from *Piezo2*^{CKO} mice was almost

absent compared to controls with all velocity ramps tested (Fig. 3b–d); in addition, these fibres displayed markedly reduced static responses. The prominent deficit in slowly adapting Aβ fibres described here is consistent with the hypothesis that Piezo2 in both Merkel cells and sensory afferents contributes to SAM I firing activity^{4,5}. Rapidly adapting mechanoreceptors that innervate hair follicles also increase their firing rates with increasing stimulus speed¹⁹. This velocity sensitivity was almost absent in the remaining rapidly adapting mechanoreceptors encountered in *Piezo2*^{CKO} mutant mice (Fig. 3e).

Examination of the mechanical thresholds and stimulus response functions of Aδ and C fibres revealed that Aδ mechanonociceptors exhibited elevated thresholds (Fig. 3f and Extended Data Fig. 5b), but the mechanosensitivity of C-fibre nociceptors were unaffected in *Piezo2*^{CKO} mutant mice (Fig. 3f and Extended Data Fig. 5c). Mechanoreceptors called D-hairs with Aδ-fibre axons were encountered less frequently in *Piezo2*^{CKO} mutant mice compared to controls but this did not reach statistical significance (Fisher's exact test, *P* > 0.1) (Extended Data Table 2), but the four fibres encountered showed stimulus response properties indistinguishable from those in controls (Extended Data Fig. 5d). Most C fibres are polymodal and respond to chemical and thermal stimuli in addition to mechanical pressure. Using a Peltier device, the thermal sensitivity of C fibres was tested, and both heat threshold for the first spike and the mean firing rates to the standard heat stimulus (ramp 1 °C per s, 32–48 °C) were unchanged in *Piezo2*^{CKO} mutant mice compared to controls (Fig. 3g and Extended Data Fig. 5e). Noxious cold sensitivity of *Piezo2*^{CKO} mutant C fibres was also unaffected (data not shown). We and others have recorded C fibres with very low mechanical thresholds in hairy skin, but their rarity precluded a detailed analysis as we did not encounter any in our control sample (Extended Data Table 2).

Our electrophysiological findings from hairy skin demonstrate that mechanosensitivity of DRG neurons to low threshold forces is predominantly mediated by Piezo2; however, the mechanosensitivity of most nociceptors is largely unaffected in *Piezo2*^{CKO} mice. Any remaining mechanosensitivity in *Piezo2*^{CKO} Aβ fibres could be due to incomplete Piezo2 ablation in *Piezo2*^{CKO} mice or the presence of as-yet-unidentified mechanically activated ion channels. Our finding of an elevation in the mean mechanical threshold for activation of Aδ mechanonociceptors suggests that Piezo2 may contribute partly to mechanosensitivity of these neurons, a finding consistent with recent data indicating that single sensory neurons often possess multiple mechanosensitive currents²⁰.

We next performed a battery of behavioural assays to test innocuous and noxious mechanical and thermal sensitivities of *Piezo2*^{CKO} mice. We tested for the ability of *Piezo2*^{CKO} mice to detect von Frey filaments of varying forces applied to their hind paws²¹. Whereas *Piezo2*^{WT} mice showed a linear increase in detection of von Frey filaments from 1.0 g to 5.5 g, *Piezo2*^{CKO} mice were severely impaired in their ability to respond to forces below 4.0 g (Fig. 4a, see Methods for an explanation of the units used). Notably, *Piezo2*^{CKO} mice showed no differences in sensing the higher forces, highlighting the role of Piezo2 in a specific range of mechanical stimuli and indicating that *Piezo2*^{CKO} mice were able to perform motor functions. We then used the cotton swab assay where a sweeping motion of cotton underneath the mouse paw led to consistent withdrawal responses in *Piezo2*^{WT} mice, whereas *Piezo2*^{CKO} mice showed markedly reduced responses to the cotton swab stimulus (Fig. 4b)²². We also built a novel two-choice preference assay to evaluate innocuous mechanosensation (Extended Data Fig. 6a, b) and found that C57BL/6J mice consistently avoided the mechanically active side and spent time on the inactive side (Extended Data Fig. 6c)²³. *Piezo2*^{WT} mice behaved similarly to C57BL/6J mice and showed robust avoidance behaviour to this stimulus, whereas the *Piezo2*^{CKO} mice showed no preference and spent equal time in the mechanically active and inactive sides (Fig. 4c).

In the assays described above, the paws of the mice receive the majority of the mechanical stimulation. We therefore tested the response of *Piezo2*^{CKO} mice to a mechanical stimulus given to hairy skin. We used a modified version of the tape response assay to monitor the response of mice to adhesive tape attached to their back²⁴. *Piezo2*^{WT} mice showed

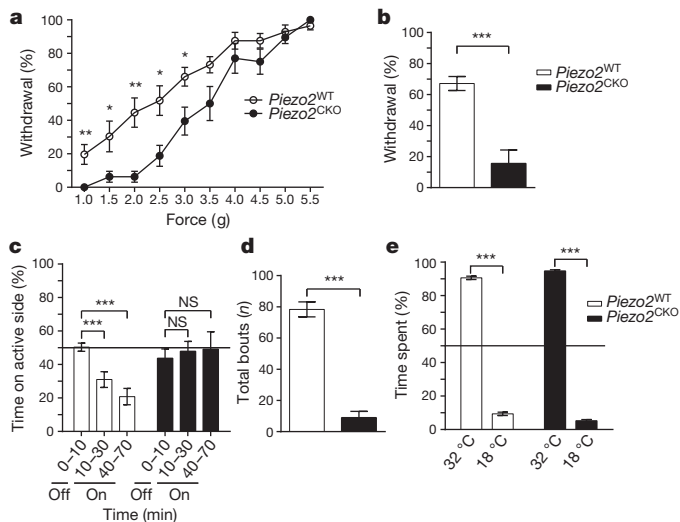


Figure 4 | *Piezo2*^{CKO} mice show profound and specific defects in innocuous touch sensation in multiple behavioural assays. **a**, Per cent response to varying forces (see Methods for an explanation of the units) of von Frey filaments in *Piezo2*^{WT} ($n = 14$) and *Piezo2*^{CKO} mice ($n = 12$). **b**, Per cent response to a sweep of a cotton swab on the hind paw in *Piezo2*^{WT} ($n = 14$) and *Piezo2*^{CKO} mice ($n = 9$). Of note, 6 out of 9 *Piezo2*^{CKO} mice showed zero responses to the cotton stimulus. **c**, Measurement of time spent in a behavioural assay where mice are placed in an apparatus consisting of two platforms, with one side mechanically active and the other side kept still. For the first 10 min, no stimulation is given to either side ('off') to allow for acclimation. Mechanical stimulation is then provided to one side for 1 hour and per cent time spent on the two sides is plotted, *Piezo2*^{WT} ($n = 15$) and *Piezo2*^{CKO} mice ($n = 11$). **d**, Number of bouts in response to a 3 cm piece of adhesive tape affixed to the back of a mouse over a 5 min period in *Piezo2*^{WT} ($n = 14$) and *Piezo2*^{CKO} mice ($n = 10$). **e**, Percentage of time spent on either side of a two-choice assay with one zone set at 18 °C and one zone set at 32 °C in *Piezo2*^{WT} ($n = 13$) and *Piezo2*^{CKO} mice ($n = 10$). Error bars represent s.e.m., all experiments performed with at least four separate cohorts of both male and female littermate control mice, * $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$, Mann-Whitney non-parametric analysis.

clear responses to the tape and made ~80 attempts to remove the tape over the course of 5 min. *Piezo2*^{CKO} mice barely responded to this stimulus (Fig. 4d). These data demonstrate that *Piezo2*^{CKO} mice have a dramatic inability to sense innocuous touch. Notably, when Piezo2 is deleted in epidermal Merkel cells but not in DRG neurons using the *K14-cre* line, these mice had a slight defect in response to von Frey filaments, but no defects in the other three innocuous touch behavioural assays described here (data not shown)⁴. Importantly, to confirm that this behaviour is not due to Piezo2 deletion in non-sensory tissues, we performed immunostaining using the *Piezo2-GFP* reporter mouse line and were unable to detect Piezo2 expression in motor neurons, cerebellum or skeletal muscle (data not shown). Therefore, the considerably more pronounced behavioural deficits resulting from Piezo2 deletion in *Advil-creERT2* mice emphasize the primary role of sensory neurons in touch sensation.

We then investigated whether *Piezo2*^{CKO} mice are able to respond to thermal stimuli and noxious mechanical forces. *Piezo2*^{CKO} mice showed no differences compared to littermate controls in innocuous temperature sensation, demonstrating that the deficits in mechanical insensitivity are not due to general health issues (Fig. 4e)²³. To determine whether *Piezo2*^{CKO} mice could respond to noxious mechanical stimuli, we first used a ramping von Frey protocol and found that *Piezo2*^{CKO} mice showed no differences in response threshold compared to *Piezo2*^{WT} mice (Extended Data Fig. 7a). These results are in line with the static force von Frey assay where the major deficit was observed only for low-threshold stimuli. *Piezo2*^{CKO} mice also showed similar responses to *Piezo2*^{WT} mice in noxious force sensing assays such as tail clip and the Randall–Selitto

test (Extended Data Fig. 7b, c). Furthermore, we observed no differences to noxious temperature sensitivity between control and *Piezo2*^{CKO} mice (Extended Data Fig. 7d). Recent reports have suggested a role for Piezo2 in mechanical allodynia^{12,25}. Although Piezo2 activity is sensitized in response to inflammatory signals *in vitro*, we did not observe any behavioural deficits in mechanical nociception or mechanical allodynia in response to inflammatory mediators in *Piezo2*^{CKO} mice (Extended Data Fig. 7e, f)¹⁰. Future studies are needed to explore Piezo2 mechanical sensitivity in response to nerve injury.

Ion channels responsible for detecting mechanical stimuli relevant for somatosensation have long been sought, and our results demonstrate that Piezo2 is the major mechanotransducer required for touch sensation in mammals. The loss of the MEC-2-related protein STOML3 also leads to loss of mechanoreceptor sensitivity^{15,20} and recent data show that this protein is a powerful positive modulator of Piezo channel mechanosensitivity. We show that Piezo2-dependent rapidly adapting, mechanically activating currents *in vitro* account for the mechanosensitivity of the majority of LTMRs without affecting the function of nociceptors. It is interesting that a Piezo2-mediated rapidly adapting current *in vitro* can account for both rapidly and slowly adapting fibre responses *in vivo*. At least in Merkel cells, the small remaining Piezo2-mediated current during long-lasting stimulations is sufficient to produce large sustained depolarization, probably due to the high membrane resistance of these cells, which leads to slowly adapting responses in sensory afferents^{4,5}. Our electrophysiological results are unambiguously supported by behavioural studies that show an almost complete deficit in touch sensation without affecting pain sensation in Piezo2-deficient mice, and predict the existence of other mechanotransducers relevant for pain sensation. Piezo1, a related mechanosensitive ion channel required for vascular development, is not expressed at significantly high levels in sensory neurons to be a likely candidate (Extended Data Fig. 8a, b)²⁶. Ion channels that sense cell volume, such as the recently identified SWELL1, might also play a role in nociception^{27,28}. Furthermore, the as-yet-unknown cationic channels that account for the intermediately and slowly adapting, mechanically activated currents in cultured DRGs (Fig. 2d) are likely to be involved in nociception. Different mechanotransducers also account for innocuous and noxious mechanosensation in *Drosophila*². Notably, *Drosophila* Piezo is required for mechanical nociception, and not for gentle touch sensation²⁹. Regardless, our findings highlight the role of Piezo proteins as evolutionarily conserved mechanically activated ion channels involved in somatosensory mechanotransduction from flies to mammals^{29,30}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 July; accepted 17 October 2014.

1. Abraira, V. E. & Ginty, D. D. The sensory neurons of touch. *Neuron* **79**, 618–639 (2013).
2. Árnadóttir, J. & Chalfie, M. Eukaryotic mechanosensitive channels. *Annu. Rev. Biophys.* **39**, 111–137 (2010).
3. Coste, B. *et al.* Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. *Science* **330**, 55–60 (2010).
4. Woo, S.-H. *et al.* Piezo2 is required for Merkel-cell mechanotransduction. *Nature* **509**, 622–626 (2014).
5. Maksimovic, S. *et al.* Epidermal Merkel cells are mechanosensory cells that tune mammalian touch receptors. *Nature* **509**, 617–621 (2014).
6. Ikeda, R. *et al.* Merkel cells transduce and encode tactile stimuli to drive A β -afferent impulses. *Cell* **157**, 664–675 (2014).
7. Maricich, S. M., Morrison, K. M., Mathes, E. L. & Brewer, B. M. Rodents rely on Merkel cells for texture discrimination tasks. *J. Neurosci.* **32**, 3296–3300 (2012).
8. Lechner, S. G. & Lewin, G. R. Hair sensation. *Physiology (Bethesda)* **28**, 142–150 (2013).
9. Maksimovic, S., Baba, Y. & Lumpkin, E. A. Neurotransmitters and synaptic components in the Merkel cell-neurite complex, a gentle-touch receptor. *Ann. NY Acad. Sci.* **1279**, 13–21 (2013).
10. Dubin, A. E. *et al.* Inflammatory signals enhance piezo2-mediated mechanosensitive currents. *Cell Rep.* **2**, 511–517 (2012).
11. Häberle, H. *et al.* Molecular profiling reveals synaptic release machinery in Merkel cells. *Proc. Natl Acad. Sci. USA* **101**, 14503–14508 (2004).

12. Lou, S., Duan, B., Vong, L., Lowell, B. B. & Ma, Q. Runx1 controls terminal morphology and mechanosensitivity of VGLUT3-expressing C-mechanoreceptors. *J. Neurosci.* **33**, 870–882 (2013).
13. McCarter, G. C., Reichling, D. B. & Levine, J. D. Mechanical transduction by rat dorsal root ganglion neurons *in vitro*. *Neurosci. Lett.* **273**, 179–182 (1999).
14. Hu, J. & Lewin, G. R. Mechanosensitive currents in the neurites of cultured mouse sensory neurones. *J. Physiol.* **577**, 815–828 (2006).
15. Wetzel, C. *et al.* A stomatin-domain protein essential for touch sensation in the mouse. *Nature* **445**, 206–209 (2007).
16. Moshourab, R. A., Wetzel, C., Martinez-Salgado, C. & Lewin, G. R. Stomatin-domain protein interactions with acid-sensing ion channels modulate nociceptor mechanosensitivity. *J. Physiol.* **591**, 5555–5574 (2013).
17. Maricich, S. M. *et al.* Merkel cells are essential for light-touch responses. *Science* **324**, 1580–1582 (2009).
18. Wellnitz, S. A., Lesniak, D. R., Gerling, G. J. & Lumpkin, E. A. The regularity of sustained firing reveals two populations of slowly adapting touch receptors in mouse hairy skin. *J. Neurophysiol.* **103**, 3378–3388 (2010).
19. Rugiero, F., Drew, L. J. & Wood, J. N. Kinetic properties of mechanically activated currents in spinal sensory neurons. *J. Physiol. (Lond.)* **588**, 301–314 (2010).
20. Poole, K., Herget, R., Lapatsina, L., Ngo, H. D. & Lewin, G. R. Tuning Piezo ion channels to detect molecular-scale movements relevant for fine touch. *Nature Commun.* **5**, 3520 (2014).
21. Petrus, M. *et al.* A role of TRPA1 in mechanical hyperalgesia is revealed by pharmacological inhibition. *Mol. Pain* **3**, 40 (2007).
22. Garrison, S. R., Dietrich, A. & Stucky, C. L. TRPC1 contributes to light-touch sensation and mechanical responses in low-threshold cutaneous sensory neurons. *J. Neurophysiol.* **107**, 913–922 (2012).
23. Dhaka, A. *et al.* TRPM8 is required for cold sensation in mice. *Neuron* **54**, 371–378 (2007).
24. Bradbury, E. J. *et al.* Chondroitinase ABC promotes functional recovery after spinal cord injury. *Nature* **416**, 636–640 (2002).
25. Eijkelkamp, N. *et al.* A role for Piezo2 in EPAC1-dependent mechanical allodynia. *Nature Commun.* **4**, 1682 (2013).
26. Ranade, S. S. *et al.* Piezo1, a mechanically activated ion channel, is required for vascular development in mice. *Proc. Natl Acad. Sci. USA* **111**, 10347–10352 (2014).
27. Voss, F. K. *et al.* Identification of LRRC8 heteromers as an essential component of the volume-regulated anion channel VRAC. *Science* **344**, 634–638 (2014).
28. Qiu, Z. *et al.* SWELL1, a plasma membrane protein, is an essential component of volume-regulated anion channel. *Cell* **157**, 447–458 (2014).
29. Kim, S. E., Coste, B., Chadha, A., Cook, B. & Patapoutian, A. The role of *Drosophila* Piezo in mechanical nociception. *Nature* **483**, 209–212 (2012).
30. Faucher, A., Nargeot, J., Mangoni, M. E. & Jopling, C. *piezo2b* regulates vertebrate light touch response. *J. Neurosci.* **33**, 17089–17094 (2013).

Acknowledgements We acknowledge T. Goode for help in behavioural analysis. We also thank R. Moran and T. Johnson for assistance with histology, K. Spencer for imaging analysis and M. Braunschweig for technical assistance. S.S.R. was funded by a pre-doctoral fellowship from the California Institute of Regenerative Medicine. R.A.M. was supported by Clinical Research Fellowship from the Max Delbrück Center (MDC). G.R.L.'s laboratory was supported by senior European Research Council grant (project 294678) and a grant from the Deutsche Forschungsgemeinschaft (SFB958 project A9). A.P. is a Howard Hughes Medical Institute Investigator. This study was partly funded by NIH grant R01 DE022358 to A.P.

Author Contributions S.S.R. and A.P. designed experiments and wrote the manuscript along with G.R.L. S.S.R., J.N.W. and Z.Q. generated transgenic lines used in this study. S.-H.W. performed all immunostaining experiments. A.E.D. conducted electrophysiology on cultured DRG neurons and J.M. isolated and cultured cells. R.A.M., C.W., V.B. and G.R.L. performed skin nerve electrophysiology and analysed the data. S.S.R., M.P., A.G.F. and K.R. performed behavioural assays on mice. B.C., J.M. and A.J.W. developed the novel two-choice mechanosensory instrument.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.P. (ardem@scripps.edu).

METHODS

All animal procedures were approved by TSRI Institutional Animal Care and Use Committees or were approved by the animal welfare office of federal state of Berlin.

Immunofluorescence. Dorsal root ganglia neurons (DRG), dorsal skin and footpads were collected from 7-week-old mice. Tissues were briefly fixed in 4% PFA and were processed as previously described⁴. Fluorescent images were captured using a Nikon confocal microscope C2. Antibodies used for immunofluorescence staining were previously reported with the addition of thymidine hydroxylase (Millipore) and CGRP (Abcam)⁴. Immunostaining images from Ai9 reporter mice crossed with the *Adil-creERT2* line were visualized by tdTomato epifluorescence^{31,32}.

Cell culture and small interfering RNA of DRG neurons. DRG cultures and transfection of short interfering RNA (siRNA) (for Piezo1 experiments) were performed exactly as previously described³. Reagents: mouse Piezo1 siRNA smart pool Dharmacon (ON-TARGETplus mouse Piezo1 (234839) siRNA SMARTpool; catalog no. L-061455-00-0005); scramble siRNA controls, Qiagen Allstars Negative Control (SI03650318).

In situ hybridization. DRG neurons from C57BL/6J mice were harvested from all levels and fixed in 10% formalin overnight. The DRG neurons were then dehydrated through an ethanol series/xylene and embedded in paraffin. Ten-micrometre sections were cut and *in situ* hybridization was carried out using the RNAscope assay (Advanced Cell Diagnostics; Hayward, CA) according to the manufacturer's instructions. Development of signal was performed using the RNAscope 2.0 HD brown detection kit. Probes for mPiezo1 (cat. number: 400181) and mPiezo2 (cat. number: 400191) were purchased from Advanced Cell Diagnostics (Hayward, CA). Slides were mounted with Cytoseal and imaged under a bright field microscope. Slides were scanned and images captured using the Nanozoomer 2.0 HT (Hamamatsu).

Electrophysiology. Whole-cell patch clamp recordings were performed as described using standard methods to achieve low access resistance^{4,10}. Cells were maintained at 22–24 °C in: 127 mM NaCl, 3 mM KCl, 1 mM MgCl₂, 2.5 mM CaCl₂, 10 mM dextrose, 10 mM HEPES (pH 7.3). Electrodes had resistances of 4.6 ± 0.7 M Ω ($n = 117$) when filled with gluconate-based low chloride intracellular solution: 125 mM K-gluconate, 7 mM KCl, 1 mM CaCl₂, 1 mM MgCl₂, 10 mM HEPES, 1 mM tetraK-BAPTA, 4 mM Mg-ATP, 0.5 mM Na-GTP (pH 7.3 with KOH). Neurons with diameters of 14 to 41 μ m were tested for mechanosensitivity using a fire-polished glass probe; the average cell diameter was similar among genotypes (23.1 ± 0.8 μ m ($n = 57$) and 22.7 ± 0.7 μ m ($n = 62$), respectively). The probe displacement was advanced in increments of 0.5 or 1 μ m. All data were analysed as previously described^{3,10}.

Ex vivo skin nerve preparation. The skin nerve preparation was used essentially as previously described to record from single primary afferents¹⁶. Mice were euthanized by CO₂ inhalation for 2–4 min followed by cervical dislocation. The saphenous nerve and the shaved skin of the hind limb were dissected free and placed in an organ bath at 32 °C. The chamber was perfused with a synthetic interstitial fluid (SIF buffer): 123 mM NaCl, 3.5 mM KCl, 0.7 mM MgSO₄, 1.7 mM NaH₂PO₄, 2.0 mM CaCl₂, 9.5 mM sodium gluconate, 5.5 mM glucose, 7.5 mM sucrose, and 10 mM HEPES (pH 7.4). The skin was placed with the corium side up in the organ bath. The saphenous nerve was placed in an adjacent chamber on a mirror to aid and under microscopy fine filaments were teased from the nerve and placed on the recording electrode. Electrical isolation was achieved with mineral oil.

For the electrical search protocol a microelectrode (0.5–1 M Ω) was manoeuvred gently to contact the epineurium of saphenous nerve and deliver electrical stimulations at 1 s intervals with square pulses of 50–500 ms duration. In most filaments 5–7 single units were encountered. The electrical nerve stimulation was done at up to 3 sites (corresponding to major branching points of the saphenous nerve) to trace electrically identified units to their receptive fields. Mechanical sensitivity of single units was tested by mechanical stimulation of their receptive field with a glass rod; units not responding to mechanical probing were designated as mechano-insensitive. Based on the conduction velocity, these units were categorized as mechanically insensitive A β , A δ and C fibres.

Mechanically sensitive units were characterized by conduction velocity (calculated by dividing conduction distance over electrical latency for the spike) and stimulus response function was obtained by a standardized mechanical stimulation protocol. The stimulating probe, equipped with a force transducer (Kleindiek, Reutlingen, Germany), was placed onto a spot within the receptive field where the most reliable responses could be obtained. The probe was a stainless steel metal rod and the diameter of the flat circular contact area was 0.8 mm. Two mechanical stimulation protocols were employed. The first stimulation protocol was applied to SAM, RAM and D-hair receptors. A piezo actuator (Physik Instrumente) was used to deliver five ramp and hold stimuli (90 μ m displacement, 2 s hold-phase and 30 s interstimulation interval) with increasing ramp velocities (0.075, 0.15, 0.45, 1.5 and 15 mm s⁻¹). The second stimulation protocol was applied to A δ mechanonociceptors and C fibres and consisted of an ascending series (6 ramp and hold stimuli) of 10 s displacement stimuli (32–1,024 μ m) at 30-s intervals, sent as a pre-programmed

series of commands to a computer-controlled nanomotor (Kleindiek, Reutlingen, Germany). Mechanical thresholds for nociceptors were measured by reading off the force (obtained from the attached force transducer) at which the first spike was obtained. The signal driving the movement of the linear motor and raw electrophysiological data were collected with a Powerlab 4.0 system (ADInstruments) and spikes were discriminated off-line with the spike histogram extension of the software. Data were obtained from between 6 and 20 skin nerve preparations per genotype.

Tamoxifen injections. Fifteen milligrams of tamoxifen (Sigma) was dissolved into 1 ml of 100% corn oil and made fresh daily before use. *Piezo2*^{WT} and *Piezo2*^{CKO} mice both received tamoxifen injections (intraperitoneal) at 150 mg kg⁻¹ for 5 consecutive days. Due to variability in tamoxifen injections and the observation of incomplete deletion in *Piezo2*^{fl/fl} mice, *Piezo2*^{fl/-} mice, containing one floxed allele and one null allele, were used for analysis. *Piezo2*^{CKO} refers to mice that received tamoxifen injections. Each mouse was weighed before injection to normalize for differences in body weight. Somatosensory behavioural assays were performed on mice between 7 and 21 days after tamoxifen injections, and all mice that were tested for mechanical sensitivity were also tested for thermal sensitivity to control for any issues of general health. Before tamoxifen injections, *Piezo2*^{CKO} mice were healthy and viable with no differences compared to *Piezo2*^{WT} littermate controls in various somatosensory assays (data not shown). However, tamoxifen injections led to a visible difference in behaviour of *Piezo2*^{CKO} but not *Piezo2*^{WT} mice that started ~7 days after the last injection. *Piezo2*^{CKO} mice walked with an irregular gait, sometimes dragged their hind legs, and were occasionally uncoordinated when attempting to stand on their hind legs.

Quantitative real-time PCR analysis. qPCR analysis was performed by first isolating total RNA using TRIzol/chloroform and isopropanol precipitation from freshly isolated DRG neurons (Life Technologies). Generation of complementary DNA was achieved by reverse transcription using the QuantiTect Reverse Transcript kit (Qiagen). For qPCR, FastStart Universal probe master mix (Rox) from Roche Diagnostics was used. The reaction was run in the Eco Real-Time PCR instrument (Illumina) using 0.5 μ l of the cDNA in a 10 μ l reaction according to the manufacturer's instructions. Real-time Taqman qPCR assays were purchased from Integrated DNA Technologies with a FAM reporter dye and a non-fluorescent quencher: mouse *Piezo2* (Mm.PT.56a.32860700, Fam38b), and an internally designed mouse *Gapdh* assay (forward primer: GCACCACCAACTGCTTAG; reverse primer: GGATG CAGGGATGATGTTT; and probe: CAGAAGACTGTGGATGGCCCCCTC).

Calibrations and normalizations were done using the $2^{-\Delta\Delta C_T}$ method, where $\Delta\Delta C_T = ((C_T(\text{target gene}) - C_T(\text{reference gene})) - (C_T(\text{calibrator}) - C_T(\text{reference gene})))$. *Gapdh* was used as the reference gene for all qPCR experiments³³.

Static force von Frey. Responses to application of a von Frey filament was performed as previously described^{21,34}. Mice are acclimated in an elevated platform for 1 h, in a similar manner to that described for the cotton swab assay. A Dynamic Plantar Aesthesiometer is used to apply varying forces of a 0.5 mm von Frey filament (Ugo Basile product ID 37450). A single force is applied for one second to the hind paw and a yes/no withdrawal response is scored. Forces on the instrument are measured with units 'g' and represent grams-force, according to manufacturer's description (Ugo Basile). The initial force applied is 5.5 g and is lowered incrementally by 0.5 g, and testing ends after reaching 1.0 g. The fraction of the number of responses out of four applications per force is measured. Data are plotted as percentage of withdrawal at each force for control and test mice.

Two-choice mechanosensory assay. The novel two-choice mechanosensory assay instrument consists of two lightweight, stiff, composite platforms which are separately activated by a pair of tactile transducers (Clark Instruments). The platforms are mounted to a frame via low rate coil springs. The springs act to mechanically isolate the platforms from each other and the frame. A thin sheet of silicone rubber is fastened to the tops of the platforms and the frame. This provides a continuous surface that covers the gaps between the frame and platforms. Dividers are placed on top of the frame to create lanes. A clear plastic cover sits on all lanes to prevent mice from climbing out of the lane. Tactile transducers housed underneath the platforms are connected to a two-channel amplifier which is fed the appropriate sine wave signals via a Labview-controlled signal generator.

For behavioural analysis, individual mice were placed in one of six lanes, with a maximum of six mice tested at the same time. We optimized the output of the stimulator using multiple protocols and found that C57BL/6J mice consistently avoided the active side when using a protocol of 150 Hz (sine wave), at a peak-to-peak displacement of 33.1 μ m and a pulsing stimulus of 3 s on and 2 s off. Interestingly, mice showed decreased avoidance of the active side when the vibration stimulus was kept constant. A side at random was turned off and the other side was turned on to a pulsing stimulus. The acceleration and frequency produced were monitored using a custom-designed program (Dual Channel v.7) to ensure only one side was active. Mice were allowed to walk freely between two zones for 10 min with no stimulus and were assayed for 1 h with the pulsing protocol. EthoVision tracking system (Noldus Information Technology) was used to monitor the movement of the mice in a dark

room illuminated with infrared light. Data are plotted as the percentage of the time spent on the active zone.

Cotton swab assay. A method to assess light-touch-evoked paw withdrawal was used as described previously²². Mice were placed in an elevated acclimation chamber separated by plexiglass dividers. The floor consists of mesh-like grids that are accessible from below due to small gaps of $\sim 5 \times 5$ mm (Ugo Basile product ID 37450-277). Mice are allowed to acclimate for 1 h before testing. A cotton swab from a Q-tip is manually pulled so that it is 'puffed out' to $\sim 3\times$ the original size. A sweeping motion, rather than an upward motion, is used underneath the mouse paw and mice are assayed for a paw withdrawal response. Five sweeps are performed with at least 10 s between each sweep. The number of withdrawals out of five trials are counted and recorded as a percentage for each mouse.

Tape response assay. A modified version of the tape on paw assay was performed²⁴. Mice are allowed to acclimate in a circular plexiglass container for 5 min. A 3 cm piece of common laboratory tape was then applied gently to the back of the mouse such that it sticks to the mouse. Mice are then observed for 5 min and the total number of responses to the tape are counted. A response is scored when the mouse stops moving and bites or scratches the piece of tape or shows a visible 'wet dog shake' motion in an attempt to remove the foreign object on its back.

Ramping von Frey. Measurement of a response to a von Frey filament was performed as previously described²¹. Mice are acclimated and tested using the same Dynamic Plantar Aesthesiometer as described for static force von Frey (Ugo Basile). Instead of a static force application as described above, a ramping protocol is used that increases in force gradually from 0–50 g over the course of 20 s. The von Frey filament is applied until the mouse withdraws its hind paw. Each mouse receives four applications of the von Frey filament and the average force (threshold for withdrawal) is plotted for each mouse.

Hargreaves assay. A paw withdrawal response to infrared light is performed as previously described²¹. Mice are allowed to acclimate on a plexiglass container with a clear flat bottom for at least 60 min. The infrared light heat source is set to an intensity of 20, according to manufacturer's protocols (Ugo Basile product ID 37370). The heat source is placed under the platform where mice are acclimated and the light is applied to the hind paw. The time to paw withdrawal is measured and repeated three times for each mouse with a cutoff of 20 s. The average time to withdrawal (latency) is plotted for each mouse.

Randall–Selitto assay. Response to a noxious level of pinching force on the paw of a mouse was performed as previously described³⁴. Mice are placed into a small sling that allows the mice to be restrained and the paws to be accessible (IITC Life Sciences). A pinching force is applied to the hind paw using a Randall–Selitto device (IITC Life Sciences). Three separate pinches are applied where the force increases until reaching a 300 g cutoff. A response is scored by any visible flinch or audible vocalization. Data are plotted as the average of three responses (threshold) calculated for each mouse.

Tail clip assay. An assay to determine the time for response to an alligator clip placed on the tail of a mouse was slightly modified from ref. 35. An alligator clip commonly used for electrical wires was purchased from The Home Depot. The clips were altered by filing down the sharp teeth and then covered with a rubber casing to reduce the potential for tissue damage. We calibrated the strength of the force exerted by the clips to be 500 g. The clips were marked at the centre to ensure placement of the clip was similar for all mice tested. The clips are placed near the base of the mouse tail and then the mice are placed on a table inside a plexiglass container. A response was scored when the mice showed awareness of the clip by biting, vocalization, grasping

of tail or a jumping response. Data are plotted as the time to respond (latency) for each mouse.

Complete Freund's adjuvant with von Frey. A standard assay for inflammation due to complete Freund's adjuvant (CFA) injection in the hindpaw was performed as previously described²¹. Mice were subjected to ramping von Frey (detailed above) before injection of CFA. After baseline analysis, mice were injected in the hind paw with 10 μ l of CFA (Sigma). Mice were then rehoused in their home cages and tested by ramping von Frey the following day, 24 h post injection.

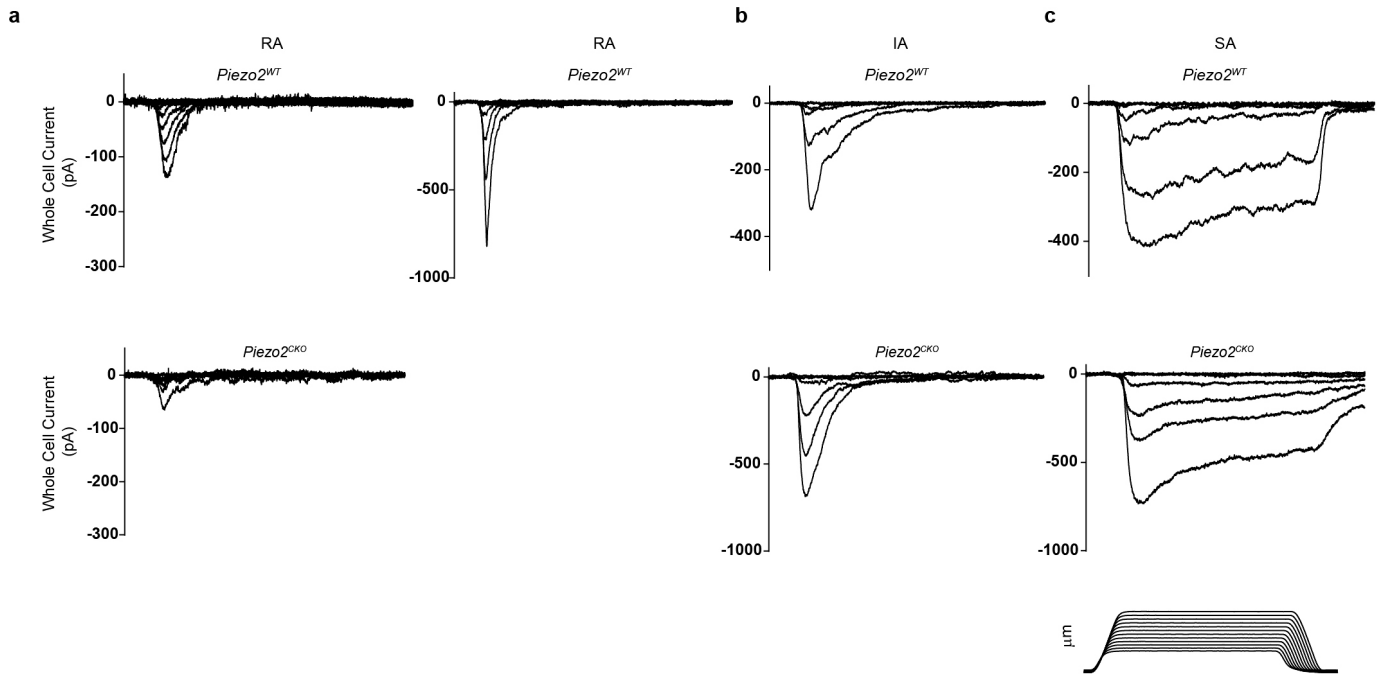
Bradykinin with von Frey. Inflammation response to bradykinin injection was performed as previously described²¹. Animals were acclimated and tested by ramping von Frey (detailed above). After baseline analysis was performed, mice were injected with 20 μ l of bradykinin (10 nM). Mice were then placed back in the acclimation chambers and tested by ramping von Frey at 5, 10 and 15 min post injection.

Two temperature choice assay. A preference assay between 32 °C and 18 °C was performed as previously described²³. Mice were placed on a platform where plexiglass dividers separated the entire platform into six lanes. The platform was divided into two zones such that one side was heated to 32 °C and the other side was cooled to 18 °C. One mouse was placed in each lane and allowed to walk freely between the two zones for 1 h. An automated tracking device (Noldus Ethovision software) monitored the movement of the mice in a dark room illuminated with infrared light. Data are plotted as the percentage of time spent on the 32 °C zone versus the 18 °C zone.

Statistics. All data analysed by statistical analyses are detailed in the figure legends. Data from electrophysiological analysis and qPCR of isolated DRG experiments were analysed by Student's *t*-test as previously described³. Electrophysiological skin nerve data were analysed using Fischer's exact test, Fig. 3a; Mann–Whitney test, Fig. 3f, g, h; and repeated measures ANOVA, Fig. 3d, e. For all behavioural experiments, the Mann–Whitney non-parametric statistical test was used.

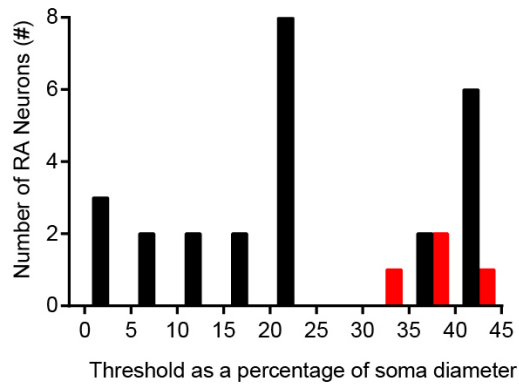
Sample size choice. The specific number of independent experiments or animal numbers used for all experiments is outlined in the corresponding figure legends. For electrophysiology and qPCR of DRG experiments, data were analysed from $n = 3$ independent experiments. For skin nerve preparations data was obtained from between 6 and 20 skin nerve preparations per genotype. All behavioural data represent results from 2–4 independent rounds of testing with multiple mice of both genotypes per cohort. All mice tested in behavioural assays were littermate controls. Male and female mice were tested between the ages of 6–12 weeks. Animals were randomly chosen for experimental subjects and all experimental procedures were performed in a blinded manner. No statistical method was used to predetermine sample size.

31. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neurosci.* **13**, 133–140 (2010).
32. Lau, J. *et al.* Temporal control of gene deletion in sensory ganglia using a tamoxifen-inducible *Advillin-Cre-ERT2* recombinase mouse. *Mol. Pain* **7**, 100 (2011).
33. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25**, 402–408 (2001).
34. Kwan, K. Y. *et al.* TRPA1 contributes to cold, mechanical, and chemical nociception but is not essential for hair-cell transduction. *Neuron* **50**, 277–289 (2006).
35. Lariviere, W. R. *et al.* Heritability of nociception. III. Genetic relationships among commonly used assays of nociception and hypersensitivity. *Pain* **97**, 75–86 (2002).

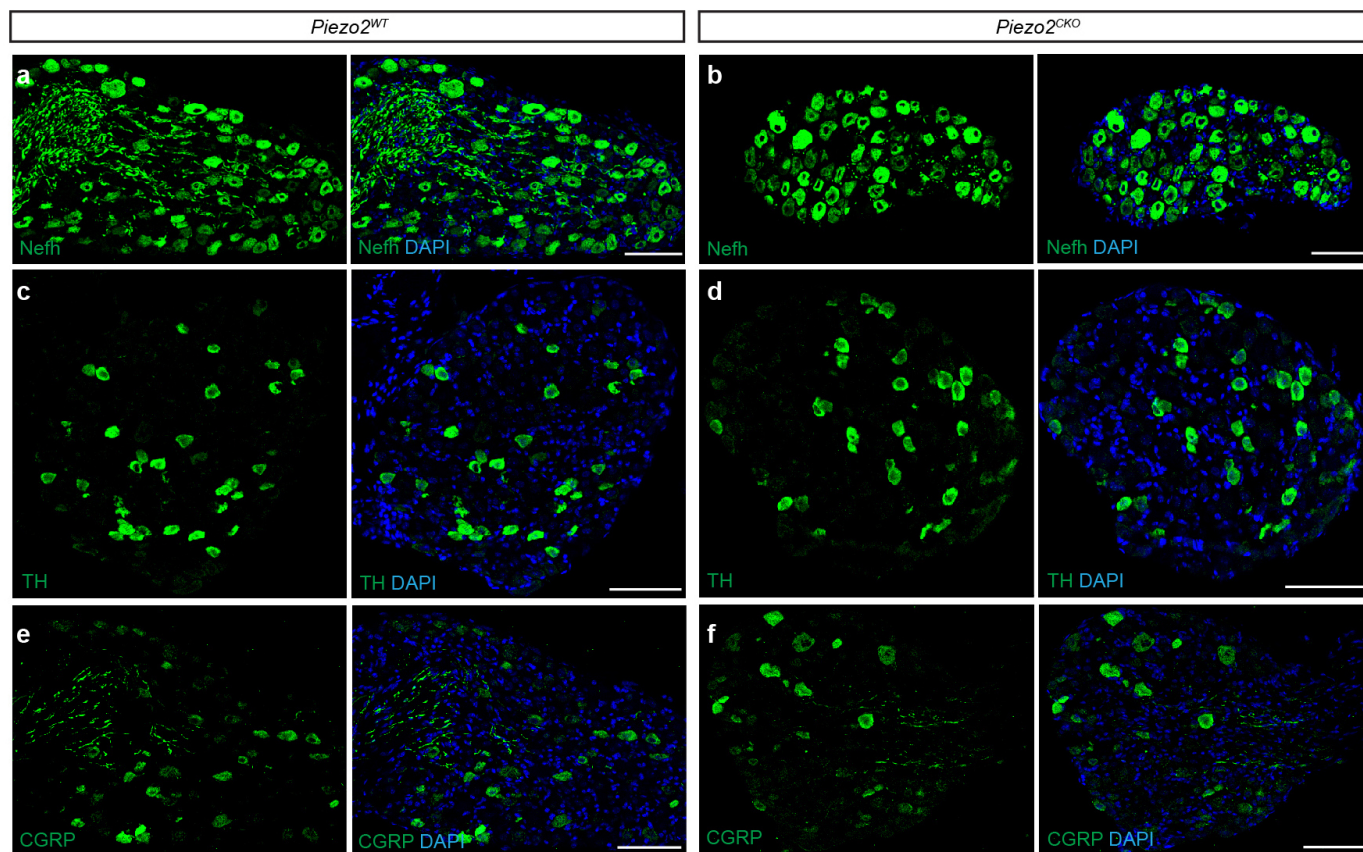


Extended Data Figure 1 | Mechanically activated currents elicited in cultured DRG neurons from *Piezo2*^{WT} and *Piezo2*^{CKO} mice by poking with a blunt probe. **a**, Representative traces of rapidly adapting (RA) currents in *Piezo2*^{WT} (top) and *Piezo2*^{CKO} (bottom). *Piezo2*^{WT} DRG neurons show characteristic rapidly adapting currents; a subpopulation can be active with apparent low thresholds (right). *Piezo2*^{CKO} mice contained a few rapidly adapting type cells but none appeared to be low-threshold mechanoreceptors. **b**, **c**, Representative traces of intermediately adapting (IA) and slowly adapting (SA) currents, respectively, with no observable differences between the two

genotypes. All data were low-pass filtered off line at 4 kHz. Action potentials were elicited by current injection in all neurons. *Piezo2*^{WT}: RA, left: 20 μ m diameter, 5 μ m apparent threshold; RA, right: 28 μ m diameter, 1 μ m apparent threshold; IA: 23 μ m diameter, 6 μ m apparent threshold; SA: 20 μ m diameter, 2 μ m apparent threshold. *Piezo2*^{CKO}: RA, left: 23 μ m diameter, 8 μ m apparent threshold; RA, right: none found; IA: 20 μ m diameter, 5.5 μ m apparent threshold; SA: 30 μ m diameter, 8 μ m apparent threshold. Lower right shows an example of the probe displacement protocol (stimulus). Results from $n = 3$ independent experiments.



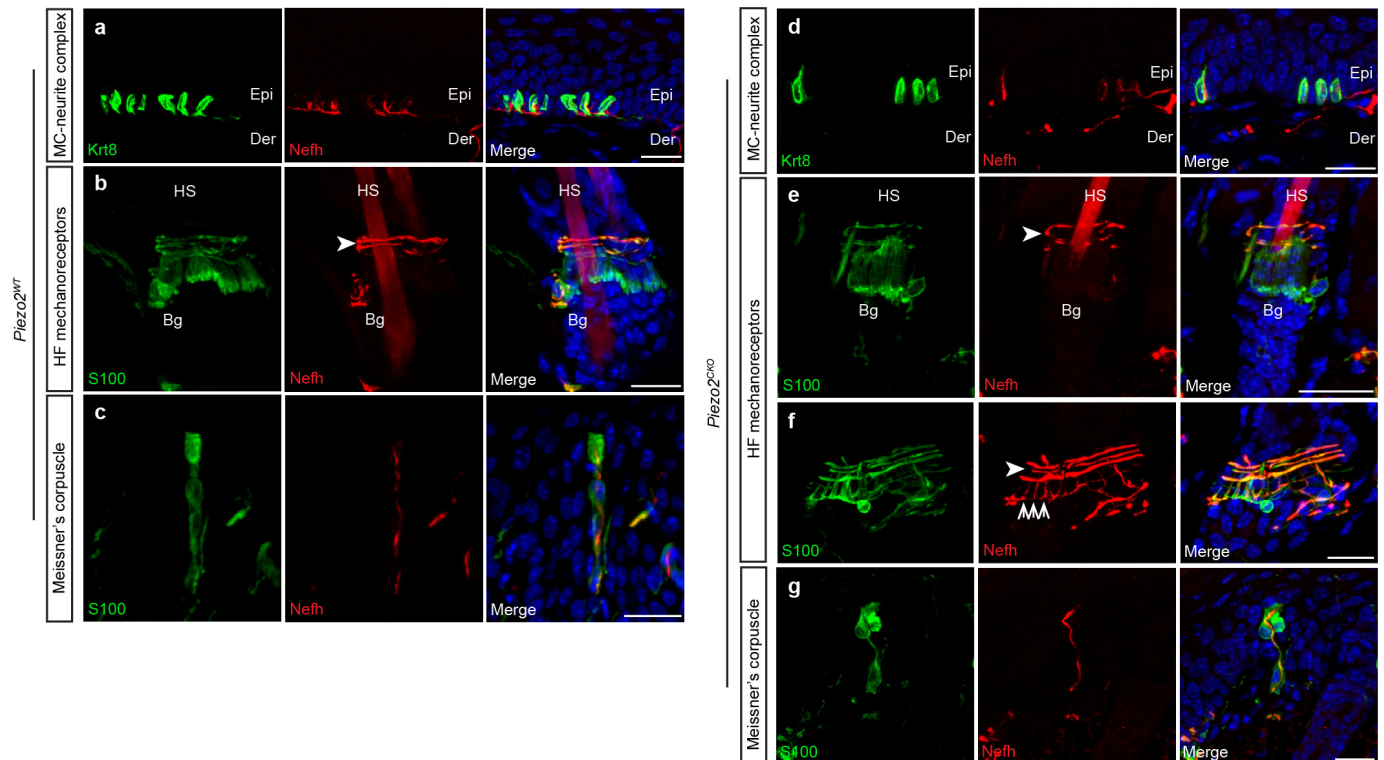
Extended Data Figure 2 | Apparent threshold analysis of *Piezo2*^{WT} and *Piezo2*^{CKO} DRG neurons. The smallest soma indentation eliciting a detectable mechanically activated response (apparent threshold) depends, in part, on the incremental distance applied (0.5 μm) and the proportional displacement in relation to the soma diameter. The apparent thresholds of all rapidly adapting responses normalized to soma diameter reveal a wide range of sensitivities of *Piezo2*^{WT} DRG neurons (black) and the high apparent threshold responses of the remaining rapidly adapting neurons in *Piezo2*^{CKO} DRG neurons (red). The lowest apparent thresholds are observed only in *Piezo2*^{WT}. Results from $n = 3$ independent experiments.



Extended Data Figure 3 | Expression of various markers of subpopulations of DRG neurons are similar in *Piezo2*^{WT} and *Piezo2*^{CKO} mice.

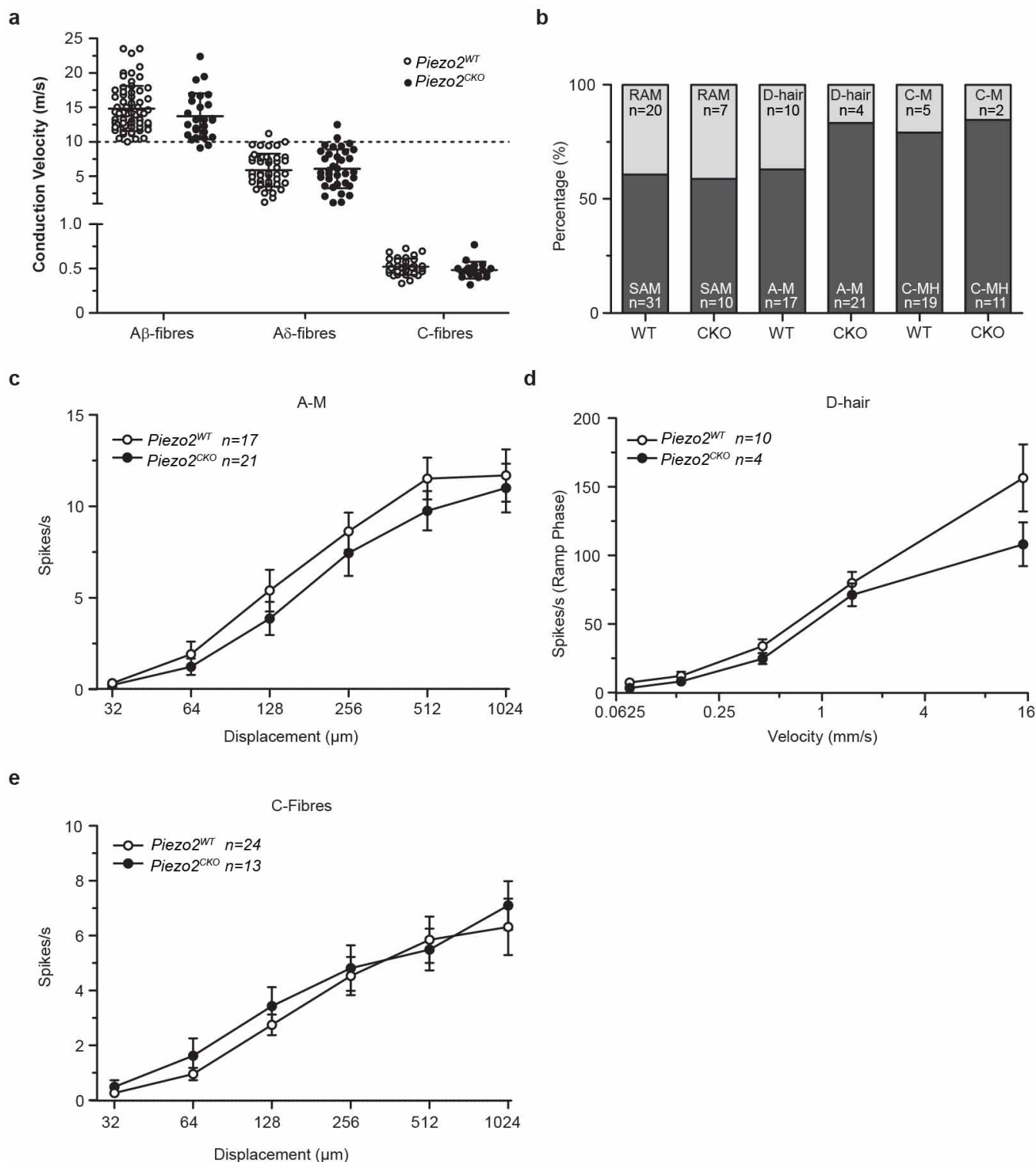
a, b, Representative images from immunofluorescence of Nefh in DRGs from *Piezo2*^{WT} (a) or *Piezo2*^{CKO} (b) mice. **c, d,** Representative image from

immunofluorescence of thymidine hydroxylase (TH) in DRGs from *Piezo2*^{WT} (c) or *Piezo2*^{CKO} (d) mice. **e, f,** Representative image from immunofluorescence of CGRP in DRGs from *Piezo2*^{WT} (e) or *Piezo2*^{CKO} (f) mice. All markers stained in green. Scale bars, 100 μ m.



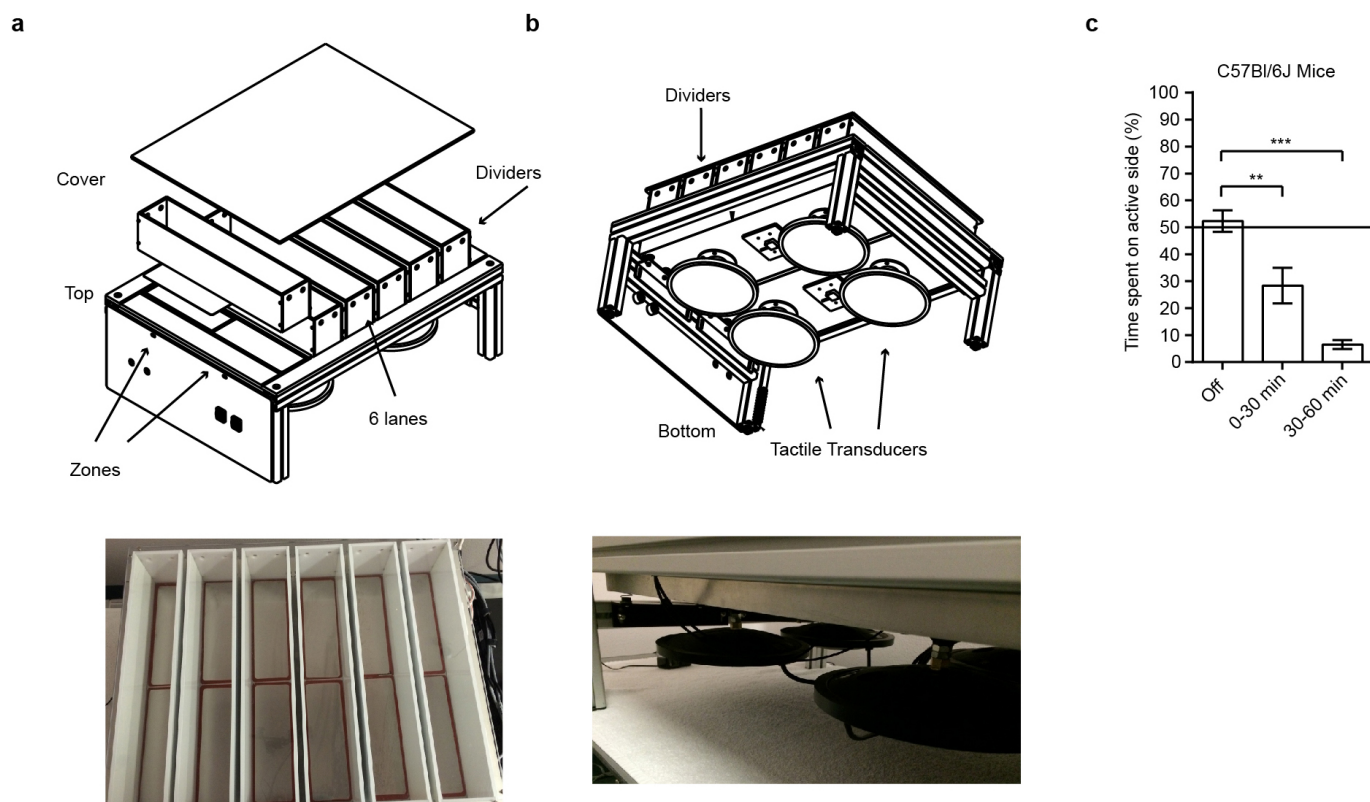
Extended Data Figure 4 | DRG innervation of skin is unaffected in *Piezo2*^{CKO} mice. **a, d**, Representative image of immunostaining of Krt8 (green) and Nefh (red) in Merkel-cell-neurite complexes in *Piezo2*^{WT} (**a**) and *Piezo2*^{CKO} (**d**) glabrous skin. **b, e, f**, Representative image of immunostaining of S100 (green) and Nefh (red) in circumferential fibres (arrowheads) and

lanceolate endings (arrows) in the hair follicle of *Piezo2*^{WT} (**b**) and *Piezo2*^{CKO} dorsal skin (**e** and **f**). **c, g**, Representative image of immunostaining of S100 (green) and Nefh (red) in Meissner's corpuscles in *Piezo2*^{WT} (**c**) and *Piezo2*^{CKO} (**g**) glabrous skin. Bg, bulge of the hair follicle; Der, dermis; Epi, epidermis; HS, hair shaft. Scale bars, 20 μ m.



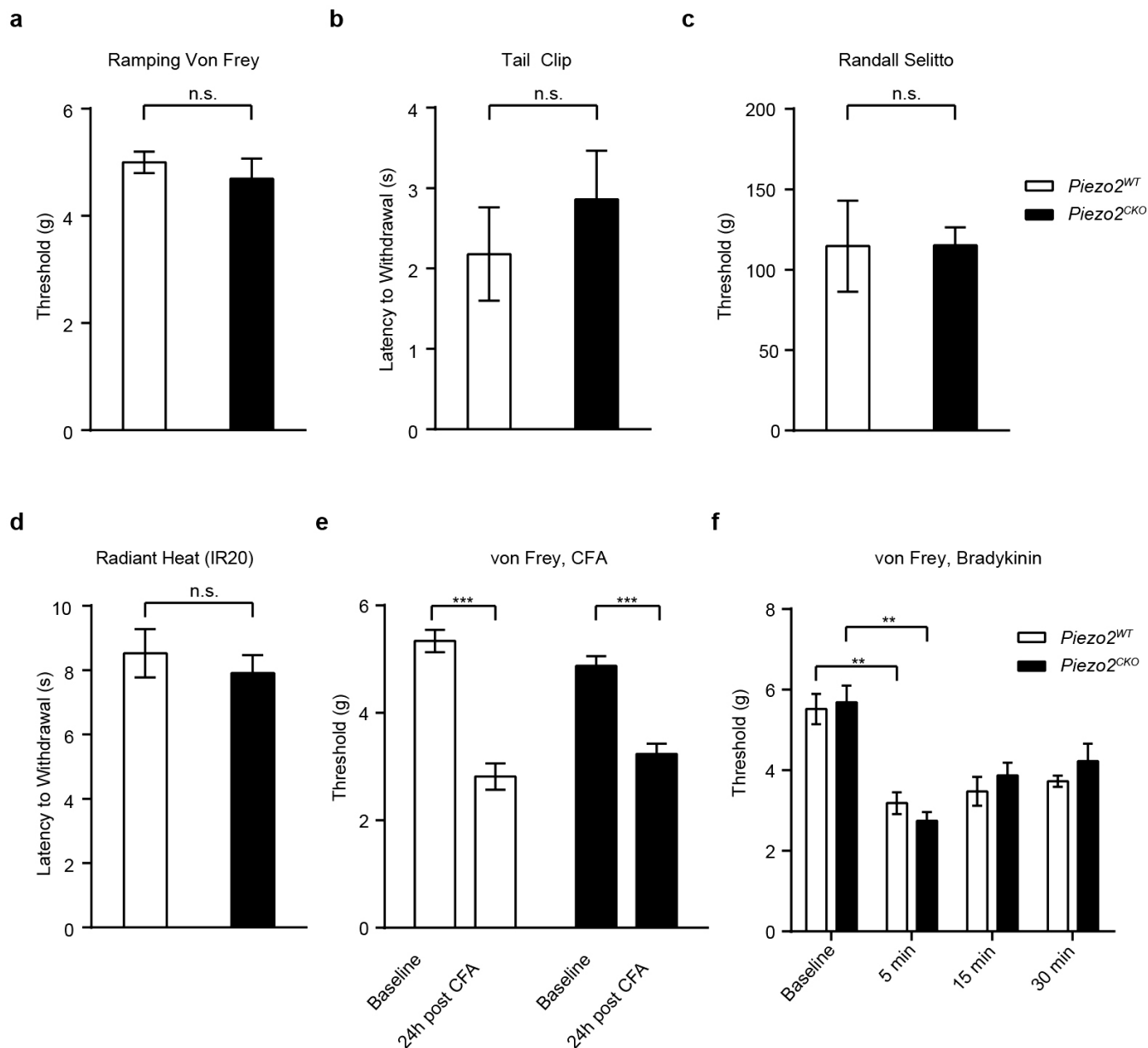
Extended Data Figure 5 | Physiological properties of nociceptors are unaffected in *Piezo2*^{CKO} mice. **a**, No change in conduction velocities of A β -, A δ - and C-fibre afferents in *Piezo2*^{CKO} compared to *Piezo2*^{WT} (Mann-Whitney test). **b**, Proportions of receptor types encountered among A β , A δ and C fibres are shown. A-M, A δ mechanonociceptor; C-M, C mechanonociceptor; C-MH, C mechano/heat receptor, responding both to noxious heat and mechanical stimuli; RAM, rapidly adapting mechanoreceptor; SAM, slowly adapting mechanoreceptor. **c**, Stimulus response properties of A δ

mechanonociceptors recorded in *Piezo2*^{CKO} compared to *Piezo2*^{WT} were not significantly different. **d**, D-hair receptors recorded from *Piezo2*^{CKO} displayed stimulus response properties that were indistinguishable from control afferents. **e**, The stimulus response properties of C fibres in *Piezo2*^{CKO} were not significantly different from C fibres recorded in control *Piezo2*^{WT} mice. Data are presented as mean \pm s.e.m., repeated measures ANOVA analysis for **c-e**.



Extended Data Figure 6 | Development of the novel two-choice mechanosensory assay. **a**, Schematic of instrument construction and image of instrument from above. **b**, Schematic of the instrument from below, with the top cover removed, and photo of tactile transducers underneath the

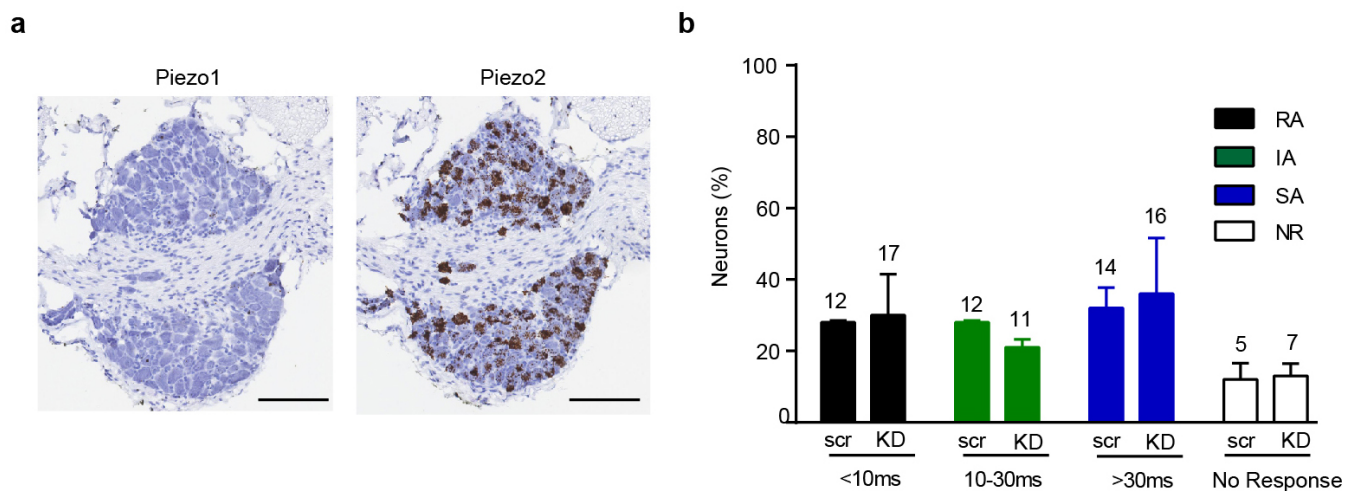
platform. **c**, Avoidance behaviour of C57BL/6J mice to the mechanically active side. Error bars represent s.e.m., $n = 12$ mice, 6 males and 6 females. $**P < 0.005$, $***P < 0.0001$, Mann-Whitney non-parametric analysis.



Extended Data Figure 7 | $Piezo2^{CKO}$ mice do not show deficits in noxious mechanical or thermal stimuli or in inflammatory pain responses.

a, Threshold for withdrawal response to a ramping protocol of von Frey stimulation from low force to high in $Piezo2^{WT}$ ($n = 9$) and $Piezo2^{CKO}$ ($n = 7$) mice. **b**, Time to response (latency) to application of a 500 g tail clip to the base of the tail in $Piezo2^{WT}$ ($n = 6$) and $Piezo2^{CKO}$ ($n = 7$) mice. **c**, Threshold for response to a Randall–Selitto pinching stimulus to the hind paw in $Piezo2^{WT}$ ($n = 5$) and $Piezo2^{CKO}$ ($n = 7$) mice. **d**, Time to withdrawal of hind paw in response to an infrared light heat source (Hargreaves assay) in $Piezo2^{WT}$

($n = 13$) and $Piezo2^{CKO}$ ($n = 9$) mice. **e**, Ramping von Frey protocol in baseline (before CFA injection) and 24 h post CFA injection in $Piezo2^{WT}$ ($n = 11$) and $Piezo2^{CKO}$ ($n = 9$) mice. **f**, Ramping von Frey protocol in baseline (before bradykinin injection) and at time points 5, 15 and 30 min post injection in $Piezo2^{WT}$ ($n = 6$) and $Piezo2^{CKO}$ ($n = 6$) mice. Error bars represent s.e.m., all experiments performed with at least two separate cohorts of both male and female mice, ** $P < 0.005$, *** $P < 0.0001$, Mann–Whitney non-parametric analysis.



Extended Data Figure 8 | Piezo1 expression and function in DRGs. **a**, *In situ* hybridization expression analysis of Piezo1 in DRG neurons relative to Piezo2. Robust expression of Piezo2 but not Piezo1 is observed, and this agrees with previously reported results using qPCR³. **b**, siRNA for Piezo1 in cultured DRG neurons does not affect the number of mechanosensitive neurons or

the ratio of rapidly, intermediately or slowly adapting currents (number of recorded neurons in each category indicated on top of bar graphs). Data from three independent preparations, not significant by Student's *t*-test. Scale bar, 100 μ m. KD, knockdown; scr, scrambled.

Extended Data Table 1 | Electrical properties of *Piezo2*^{WT} and *Piezo2*^{CKO} cultured DRG neurons

Properties	<i>Piezo2</i> ^{WT}	<i>Piezo2</i> ^{CKO}
Membrane Resistance (MΩ)	985 ± 140 (n=57)	860 ± 110 (n=62)
Access resistance, R _a (MΩ)	12.3 ± 0.5 (n=62)	12.7 ± 0.6 (n=57)
Holding current at -80mV (pA)	-272 ± 42 (n=57)	-235 ± 34 (n=62)
Resting membrane potential (mV)	-49.3 ± 0.9 (n=57)	-47.8 ± 1.1 (n=62)
Rheobase (pA)	430 ± 90 (57)	220 ± 35 (60)*
Number of AP at 2x rheobase	1.2 ± 0.1 (54)	1.2 ± 0.1 (59)
APD (msec)	2.45 ± 0.18 (57)	2.49 ± 0.15 (61)
APD range	0.4, 6	0.6, 5.2
AP overshoot (mV)	71.9 ± 1.6 (57)	70.7 ± 1.6 (61)

Values are mean ± s.e.m. (number of neurons). APD, action potential duration; AP, action potential; * $P < 0.05$, Student's *t*-test; rheobase values are not significantly different ($P=0.3$) when cells with many processes are removed from the data set (high rheobase, $n=7$ and 6 for *Piezo2*^{WT} and *Piezo2*^{CKO}, respectively).

Extended Data Table 2 | Properties of single fibres recorded from *Piezo2*^{WT} and *Piezo2*^{CKO} using the saphenous nerve *ex vivo* skin nerve preparation

a

Subclass		<i>Piezo2</i> ^{WT}	<i>Piezo2</i> ^{CKO}
A β -fibres	SAM I	22	7
	SAM II	9	3
	RAM	20	7
A δ -fibres	D-Hair	10	4
	A-M	17	21
C-fibres	C-LT	-	-
	C-M	5	2
	C-MH	15	8
	C-MHC	4	3

b

	<i>Piezo2</i> ^{WT}	<i>Piezo2</i> ^{CKO}
Mechanically Sensitive	84	58
Mechanically Insensitive	4	13

a, Mechanical search protocol. **b**, Electrical search protocol. C-LT, C fibre with low-threshold mechanosensitivity; C-M, C mechanonociceptor only responding to mechanical stimulation and not noxious cold or heat; C-MH, C mechano/heat receptor, responding both to noxious heat and mechanical stimuli; C-MHC, C mechano/heat/cold fibre responding to noxious mechanical, hot and cold stimulation; SAM II, slowly adapting mechanoreceptor type II. Note that small discrepancies between the numbers of fibres listed here and those tested with the full series of mechanical stimuli are due to technical issues preventing full characterization of the unit.

Physical mechanism for gating and mechanosensitivity of the human TRAAK K⁺ channel

Stephen G. Brohawn¹, Ernest B. Campbell¹ & Roderick MacKinnon¹

Activation of mechanosensitive ion channels by physical force underlies many physiological processes including the sensation of touch, hearing and pain^{1–5}. TRAAK (also known as KCNK4) ion channels are neurally expressed members of the two-pore domain K⁺ (K2P) channel family and are mechanosensitive⁶. They are involved in controlling mechanical and temperature nociception in mice⁷. Mechanosensitivity of TRAAK is mediated directly through the lipid bilayer—it is a membrane-tension-gated channel⁸. However, the molecular mechanism of TRAAK channel gating and mechanosensitivity is unknown. Here we present crystal structures of TRAAK in conductive and non-conductive conformations defined by the presence of permeant ions along the conduction pathway. In the non-conductive state, a lipid acyl chain accesses the channel cavity through a 5 Å-wide lateral opening in the membrane inner leaflet and physically blocks ion passage. In the conductive state, rotation of a transmembrane helix (TM4) about a central hinge seals the intramembrane opening, preventing lipid block of the cavity and permitting ion entry. Additional rotation of a membrane interacting TM2–TM3 segment, unique to mechanosensitive K2Ps, against TM4 may further stabilize the conductive conformation. Comparison of the structures reveals a biophysical explanation for TRAAK mechanosensitivity—an expansion in cross-sectional area up to 2.7 nm² in the conductive state is expected to create a membrane-tension-dependent energy difference between conformations that promotes force activation. Our results show how tension of the lipid bilayer can be harnessed to control gating and mechanosensitivity of a eukaryotic ion channel.

How physical force gates mechanosensitive ion channels in animals is unknown^{1–5}. Mechanosensitivity in the bacterial MscL and MscS ion channels is proposed to result from a membrane-tension-dependent energy difference between open and closed conformations due in part to cross-sectional area expansion with opening^{9–11}. The eukaryotic K⁺ channel TRAAK, like MscL and MscS, is gated open by membrane tension⁸. TRAAK is a member of the K2P K⁺ channel family. The architecture of K2Ps is unique among metazoan K⁺ channels because two of the four subunits that form canonical K⁺ channels are fused into a single protein chain, thus, K2Ps are dimers^{12–14}. Constraints imposed by this architecture preclude direct extension of the canonical mechanism of gating four-fold-symmetric K⁺ channels involving symmetric dilation of a helical bundle at the intracellular channel surface^{15,16}. Crystal structures of TRAAK^{12,13} revealed intramembrane openings to the lipid bilayer between subunits and a role for lipids in gating the channel has been hypothesized⁴.

Figure 1a shows the effect of stretching the membrane on TRAAK channel activity. Pressure application to inside-out patches induces a rapid and reversible increase of K⁺-selective current. The construct used for these studies removed the disordered carboxy terminus from TRAAK because it improved the resolution of X-ray crystal structures by ~0.25 Å without affecting function. We determined the structure of TRAAK purified in decyl maltoside detergent and 150 mM K⁺ in complex with Fabs from mouse monoclonal antibodies at 2.5 Å resolution (Fig. 1b, Extended Data Table 1).

We find in this structure that TRAAK is asymmetric because transmembrane helix 4 (TM4) from each subunit adopts a different conformation (Fig. 1b–e). TM4 from subunit A (TM4A) is ‘up’—it is kinked approximately halfway through the membrane around the hinge glycine conserved in K⁺ channels (G268) and packs hydrophobic residues on its extracellular-facing side against the cytoplasmic-facing side of TM2B (Fig. 1c). This creates a continuous protein surface facing the membrane that seals the channel cavity from lipids (the ~3 Å-wide gap between L276 and I151 is smaller than an acyl chain). On the other side of the channel, TM4B is ‘down’—it traverses the membrane without kinking at the hinge glycine (Fig. 1d). This creates a ~5 Å-wide intramembrane opening between TRAAK subunits that exposes the channel cavity to the inner leaflet of the lipid bilayer.

An elongated tube of electron density consistent with a ten-carbon acyl chain extends from this intramembrane opening to the central cavity underneath the selectivity filter (Fig. 1e). We cannot distinguish crystallographically whether this acyl chain is from a co-purified lipid or a detergent molecule because of the lack of strong density features outside of the channel. This suggests that TRAAK does not tightly bind the head group of the molecule occupying the cavity. Electrophysiological evidence presented below, the presence of acyl chains in analogous positions in the structure of the related K2P TWIK1 (ref. 14), the hydrophobic nature of the site (Extended Data Fig. 1) and geometric considerations (a typical acyl chain from an inner leaflet lipid can reach the site) lead us to believe that a lipid acyl chain can occupy the cavity in a membrane.

We wondered whether the presence of this acyl chain would preclude ion binding in the cavity of TRAAK. We used the anomalous dispersion of X-rays by the permeant ion Tl⁺ as a sensitive measure of ion occupancy to test this idea. A structure in which 150 mM Tl⁺ replaced K⁺ was determined at 3.0 Å resolution. The overall structure in Tl⁺ is essentially identical to the structure in K⁺ (Extended Data Table 2). The cavity is similarly exposed to the membrane on the side of TM4B and acyl chain density is observed, although the intracellular half of TM4B is apparently disordered. Anomalous density maps show five Tl⁺ ions along the conduction pathway: one above the extracellular mouth of the pore and four in the selectivity filter (Fig. 1f). No Tl⁺ is observed in the cavity site that is occupied by lipid.

We screened for crystallization conditions to capture TRAAK with both intramembrane openings sealed and identified chemical additives that promoted this conformation (Fig. 1b, g–j). The best diffracting K⁺- and Tl⁺-bound crystals contained 4% (vol/vol) polypropylene glycol or 52 mM trichloroethanol (a known TRAAK activator¹⁷) and structures were determined to 2.5 Å and 3.0 Å resolution, respectively (Extended Data Table 1). Compared to the previous structures, TM4A remains in the same up conformation (Fig. 1g). The conformation of TM4B changes by bending around the hinge G268 to pack against TM2A and seal the intramembrane opening (Fig. 1h). With both intramembrane openings sealed, elongated acyl chain density is no longer observed in the cavity of these structures. In its place, there is approximately spherical density (Fig. 1i) and the Tl⁺-containing structure has a strong (5.1σ) additional

¹Laboratory of Molecular Neurobiology and Biophysics and Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA.

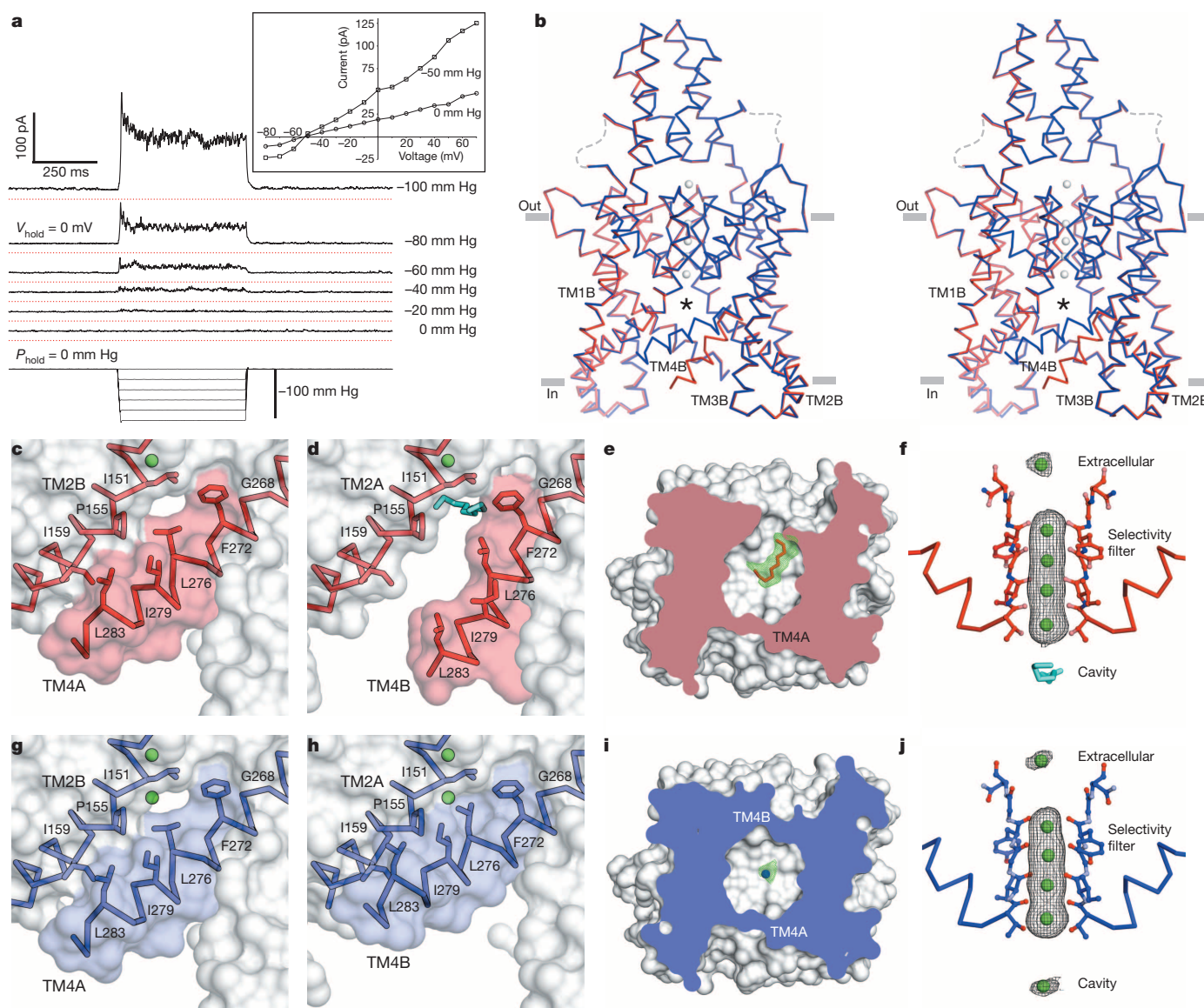


Figure 1 | Structures of TRAAK in non-conductive and conductive conformations. **a**, Current response to pressure applied to an inside-out patch from a TRAAK-expressing cell held at 0 mV in a tenfold $[K^+]$ gradient ($E_{K^+} = -59$ mV). Recordings are vertically offset with red lines at the zero current level. Inset, current–voltage relationship from the same patch before and during pressure application. **b**, Stereo view of an overlay of conductive (blue) and non-conductive (red) conformations of TRAAK in K^+ (grey). The central cavity (with ligands removed) is marked with an asterisk. **c–f**, Non-conductive TRAAK structures. **c, d**, Membrane view of the cytoplasmic half of

TM4A (**c**) and TM4B (**d**) with a cavity-bound acyl chain (cyan). Panel **d** is rotated 180° about the conduction axis from **c**. **e**, Cytoplasmic view clipped to the cavity plane with $F_o - F_c$ positive omit density at 2.5σ (green) around the acyl chain (red). **f**, Anomalous density (grey) at 3σ (extracellular ion) or 5σ (selectivity filter ions) around TI^+ (green) in the conduction pathway of non-conductive TRAAK with a cavity-bound acyl chain (cyan). **g–j**, Conductive TRAAK structures in the same views as **c–f**. The cavity-bound ion is shown in $F_o - F_c$ positive omit density (**i**, green) and anomalous density (**j**, grey) at 3σ .

anomalous peak in the cavity (Fig. 1j). Although we do not observe density for the additives that promote this conformational change, trichloroethanol functionally activates TRAAK channels in patches¹⁷.

The structures show that when TM4 is down, a lipid acyl chain can access the channel cavity through intramembrane openings and compete with ion binding. When TM4 is up, the intramembrane opening is sealed and an ion occupies the cavity site. These observations lead us to the following hypothesis: the TM4-down, lipid-blocked structure is a non-conductive TRAAK conformation, whereas the TM4-up structure is a conductive TRAAK conformation.

This hypothesis predicts that a lipid less capable of entering through the intramembrane opening would render channels more conductive. To test this, we reconstituted purified TRAAK in phosphatidylcholine with either unbranched (PC) or branched diphytanoyl (DPhPC) acyl

chains and recorded from proteoliposome patches (Extended Data Fig. 2). We observed higher levels of current in DPhPC compared to PC. This could result from either more active channels or more efficient reconstitution (that is, more channels) in DPhPC. When the patch is inflated, the mechanically activatable component is not proportionally increased as would be expected if there were simply a larger number of channels. Thus, the baseline activity of TRAAK is higher in DPhPC, consistent with the notion that the diphytanoyl chain is less able to enter the cavity and block ion conduction.

A second prediction from the model is if TM4 is trapped in the up conformation, the channel should be more conductive. To test this, we exploited the large conformational change in TM4 to design a double cysteine mutant TRAAK(I159C, R284C) that could form disulphide bridges only in the predicted conductive conformation (Fig. 2a, b). The

TRAAK(I159C, R284C) mutant was rapidly activated upon exposure to oxidizing solution (greater than tenfold within 1 min) and activation continued over several minutes to 23.8 ± 4.7 -fold (Fig. 2c, e, Extended Data Fig. 3). Wild-type channels in contrast showed only a small effect (1.67 ± 0.19 -fold activation) (Fig. 2d, e, Extended Data Fig. 3). Single cysteine mutants or three other double cysteine mutants close in space to TRAAK(I159C, R284C) showed no effect relative to wild-type (Fig. 2e). Therefore, the activation of TRAAK(I159C, R284C) in oxidizing solution depends on the three-dimensional orientation of the cysteine pair. These results suggest that the formation of a disulphide bridge between TM4 and TM2 stabilizes a conductive conformation of TRAAK.

As in the case of TRAAK reconstituted in DPhPC, the mechanically activatable component of oxidized TRAAK(I159C, R284C) is a smaller fraction of the baseline current compared to reduced TRAAK(I159C,

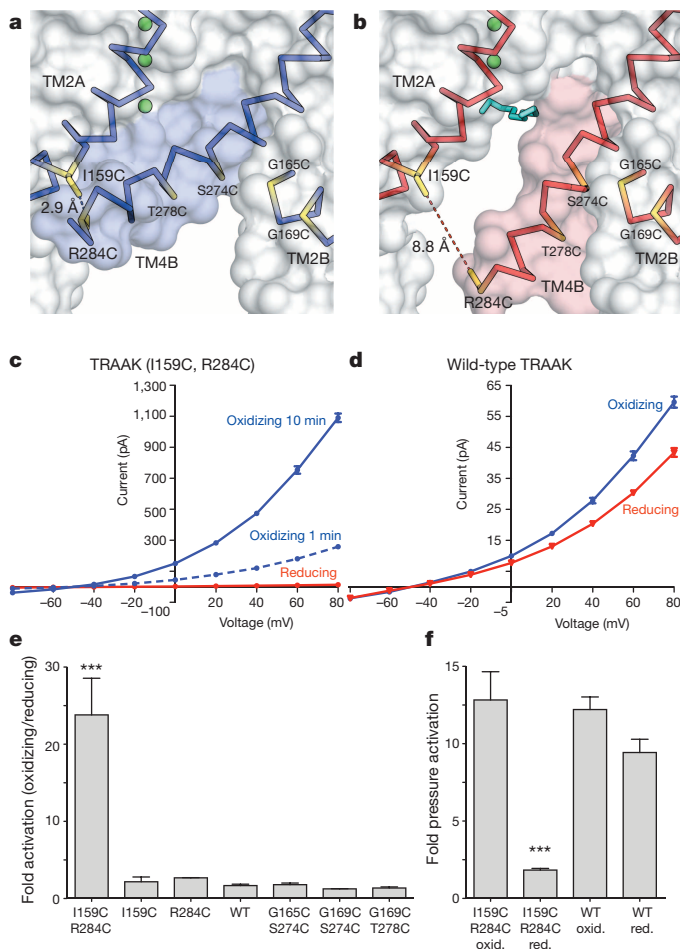


Figure 2 | Redox-dependent activation of a TRAAK double cysteine mutant between TM2 and TM4. **a, b**, Cysteine mutations (yellow) and predicted distance between sulphur atoms in TRAAK(I159C, R284C) in the conductive (**a**) and non-conductive (**b**) conformations. **c, d**, Current-voltage relationship from a representative TRAAK(I159C, R284C) (**c**) or wild-type (WT) TRAAK (**d**) patch in reducing (red) and oxidizing (blue) bath solutions.

e, Quantification of fold activation in oxidizing compared to reducing bath solutions (mean \pm s.e.m. at 0 mV from three separate experiments, TRAAK(I159C, R284C) 23.8 ± 4.7 ($n = 6$ patches); TRAAK(I159C) 2.17 ± 0.62 ($n = 3$); TRAAK(R284C) 2.68 ± 0.03 ($n = 3$); TRAAK(WT) 1.67 ± 0.19 ($n = 6$); TRAAK(G165C, S274C) 1.80 ± 0.21 ($n = 3$); TRAAK(G169C, S274C) 1.25 ± 0.02 ($n = 3$); and TRAAK(G169C, T278C) 1.37 ± 0.14 ($n = 3$); *** $P < 0.0001$, one way ANOVA). **f**, Quantification of pressure activation in oxidizing or reducing bath solutions (mean fold activation at 0 mV \pm s.e.m., TRAAK(I159C, R284C) reducing 12.83 ± 1.83 ; TRAAK(I159C, R284C) oxidizing 1.83 ± 0.11 ; TRAAK(WT) reducing 12.21 ± 0.82 ; and TRAAK(WT) oxidizing 8.43 ± 0.86 ; $n = 3$ patches each from three separate experiments, *** $P = 0.0003$, one way ANOVA).

R284C) or wild-type TRAAK (Fig. 2f, Extended Data Fig. 3). In principle, if all channels were locked open with a disulphide bridge, we should observe no further opening with mechanical perturbation. However, TRAAK open probability is low in all contexts studied (Extended Data Fig. 4). Together, the increased baseline activity and relative reduction of mechanosensitivity in these experiments are consistent with the proposed TM4-mediated gating of TRAAK.

We asked whether comparison of the structures could provide additional insight into the physical basis of TRAAK mechanosensitivity. Figure 3a shows the spectrum of conformational changes observed in TRAAK structures (Extended Data Table 3). The cytoplasmic half of TM4 is observed in one non-conductive (TM4-down, red) and two conductive (TM4-up) conformations. From the non-conductive conformation, rearrangement of the intracellular half of TM4 by a $\sim 15^\circ$ bend (around G268) and a $\sim 15^\circ$ right-handed rotation about the helical axis gives the orange TM4-up conformation. Alternatively, a $\sim 25^\circ$ bend and a $\sim 10^\circ$ rotation gives the blue, green, or yellow TM4-up conformations.

The transition from a non-conductive to conductive state can occur by movement of TM4 alone (compare the red and green structures) or with additional conformational changes (Fig. 3a). In the blue and orange structures, the linked inner extension of TM2–TM3 rotates ~ 10 – 15° towards TM4. This can only occur if TM4 is up because a rotated TM2–TM3 would sterically clash with TM4 in a down conformation (Extended Data Fig. 5). By creating a steric barrier to the non-conductive conformation of TM4, rotation of TM2–TM3 may stabilize the conductive conformation. In the blue, orange, and yellow structures, the selectivity filter to TM4 linker rearranges and the extracellular end of TM4 is laterally displaced from the conduction axis compared to the red structure. This displacement is observed only in TM4-up structures, although unlike with the TM2–TM3 rotation, there is no apparent structural constraint that necessitates the changes be coupled.

The effect of membrane tension on TRAAK activity can be observed in a simple experiment. A cell expressing TRAAK recorded in whole-cell mode gives $\sim 1,000$ pA of current (Extended Data Fig. 4a, b). An outside-out patch excised from this cell runs up over several minutes to give ~ 250 pA and applying pressure to the patch gives $\sim 1,000$ pA of current (Extended Data Fig. 4c, d). It is surprising to observe that a patch containing a small fraction of the whole cell membrane ($< 1\%$) gives currents similar in magnitude to the whole cell. A logical explanation is that membrane tension varies in these recordings. The whole-cell has near zero tension, the excised patch has low to intermediate tension owing to the adhesion of lipid to glass (~ 0.5 – 4 mN m $^{-1}$) and patch pressurization causes high tension (up to lytic values ~ 12 mN m $^{-1}$)^{5,18,19}. From these experiments we can estimate that TRAAK has very low basal activity near zero tension ($< 1\%$ of maximum activity) and is activated over a broad tension range from ~ 0.5 – 4 mN m $^{-1}$ to ~ 12 mN m $^{-1}$.

We want to understand whether the structural differences between conductive and non-conductive states can explain the mechanosensitivity of TRAAK. Membrane tension (γ) favours an expansion of protein cross-sectional area (ΔA) by an energy equal to $-\gamma\Delta A$. Plotting the change in cross-sectional area between conductive and non-conductive conformations shows a marked expansion of TRAAK with channel opening (Fig. 3b) with each conformational change described above contributing to expansion at different membrane depths. The transition to TM4-up expands the channel by up to 2.7 nm 2 in the cytoplasmic leaflet due to its elevated projection in the membrane (Fig. 3b, c). At the membrane–cytoplasm boundary, rotation of TM2–TM3 expands TRAAK by 1.5 nm 2 by sealing a crevice between TM3 and TM4. In the extracellular leaflet, displacement of the cytoplasmic half of TM4 away from the conduction axis expands TRAAK by 1 nm 2 . Over the range of tensions that activates TRAAK (~ 0.4 – 4 to 12 mN m $^{-1}$), area expansion between non-conductive and conductive TRAAK conformations gives a calculated energy difference of 0.3 – 8.1 $k_B T$ that promotes force activation.

Changes in lipid deformation may also contribute to tension activation of TRAAK. Inspection of the structures gives a qualitative appreciation of this effect. The non-conductive channel forms a distinctly angled

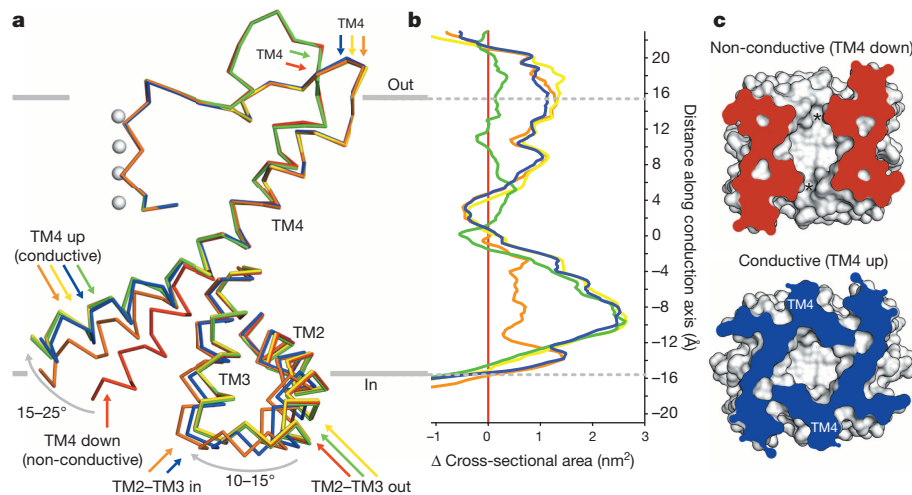


Figure 3 | Conformational changes in conductive TRAAK structures expand the channel cross-sectional area. **a**, Membrane view of an overlay of mobile elements in non-conductive (red) and conductive conformations (orange, green, yellow and blue) of TRAAK. **b**, Difference in cross-sectional area versus membrane depth calculated between each conductive conformation

cavity between subunits (bound by the inner halves of TM2, TM3 and TM4) (Figs 3 and 4). In the conductive channel this cavity becomes sealed by the elevation of TM4 and rotation of TM2–TM3 to create a flatter membrane-facing surface. The non-conductive channel is more wedge-shaped, the conductive channel more cylindrical, and these different shapes have energetic consequences²⁰. A cylinder matches the average flat plane of a membrane better than a wedge. A wedge forces the membrane to curve where it meets the surface of the channel and this curving costs a mid-plane bending free-energy penalty: the higher the tension, the higher the cost²⁰. Consequently, as membrane tension is increased the cylindrical shape becomes more energetically favoured relative to the wedge shape and tension favours opening.

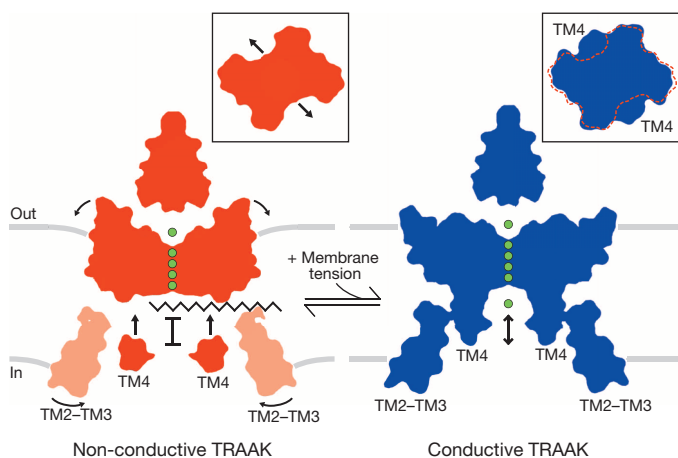


Figure 4 | Model for TRAAK gating and physical basis of mechanosensitivity. In the non-conductive conformation (red), a lipid acyl chain accesses the central cavity through intramembrane openings above TM4 to sterically block ion (green) conduction. In the conductive conformation (blue), conformational changes in TM4 seal the intramembrane openings to prevent lipid access and permit ion conduction through the channel. Conformational changes upon channel opening increase the channel cross-sectional area (insets). Rotation of TM2–TM3 towards TM4 in the cytoplasmic leaflet and displacement of the extracellular half of TM4 further expand the channel and may stabilize the conductive conformation. Area expansion and reduced mid-plane bending of the membrane (grey bars) result in a membrane-tension-dependent energy difference between conformations that explains the mechanosensitivity of TRAAK.

and the non-conductive conformation (positive values indicate area expansion upon channel opening, colours as in **a**). Depth is on the same scale as **a** with grey indicating bilayer boundaries. **c**, Cytoplasmic view of non-conductive and conductive conformations clipped at -9 Å depth. Intramembrane openings (asterisks) or TM4 positions are indicated.

Our data support a model for gating and mechanosensitivity of TRAAK shown in Fig. 4. The conformational change in the two TM4s that gates TRAAK is similar to the conformational change in four inner helices that gates canonical K⁺ channels^{15,16}. However, instead of opening a protein seal to the cytoplasm, TM4 movement in TRAAK seals intramembrane openings to prevent lipid from blocking ion conduction. It is possible that other channels might use this unprecedented gating mechanism. The closely related K2Ps TREK1 and TREK2 share the key functional and structural features of TRAAK^{8,12,21}. Predictions of area expansion and energy differences from electrophysiological studies of TREK1 agree closely with the values calculated for TRAAK here^{22,23}. Increasing the interaction of a cytoplasmic sensor region immediately following TM4 with the membrane is proposed to activate TREK channels^{24–28}, but how this sensor is physically coupled to a channel gate is unknown. Extension of the model presented here suggests a simple explanation: increasing the proximity of the intracellular end of TM4 to the membrane in TREK channels stabilizes a conductive TM4-up conformation.

Comparison of TRAAK and TWIK1 suggests an explanation for why among K2Ps only TRAAK subfamily channels are mechanosensitive⁸. TWIK1 was observed in a TM4-down conformation with intramembrane openings filled with acyl chains¹⁴. Pore dehydration of TWIK1 has been proposed to stabilize the non-conductive state and introducing polar residues in the cavity increases channel activity, consistent with a hydrophobic barrier to conduction²⁹. TRAAK has a wider cavity than TWIK1, which expands in the conductive conformation (Extended Data Fig. 1). Two structural differences appear to restrict the TM4s of TWIK1 from adopting a membrane sealing conformation. First, an amphipathic helix after TM4 in TWIK1 anchors the cytoplasmic end of the helix to the membrane. Second, TM4 in TWIK1 interacts extensively with the cytoplasmic half of TM2–TM3. In TRAAK, the amphipathic nature of TM2–TM3 results in its lateral extension in the membrane away from TM4, enabling free rotation of the inner half of TM4. If TM4 is prevented from rotating to an up conformation in TWIK1, gating would not produce the area expansion and shape changes we propose underlie the mechanosensitivity of TRAAK.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 August; accepted 30 October 2014.

1. Arnadóttir, J. & Chalfie, M. Eukaryotic mechanosensitive channels. *Ann. Rev. Biophysics* **39**, 111–137 (2010).

2. Bautista, D. M. & Lumpkin, E. A. Probing mammalian touch transduction. *J. Gen. Physiol.* **138**, 291–301 (2011).
3. Delmas, P. & Coste, B. Mechano-gated ion channels in sensory systems. *Cell* **155**, 278–284 (2013).
4. Nilius, B. & Honore, E. Sensing pressure with ion channels. *Trends Neurosci.* **35**, 477–486 (2012).
5. Anishkin, A., Loukin, S. H., Teng, J. & Kung, C. Feeling the hidden mechanical forces in lipid bilayer is an original sense. *Proc. Natl Acad. Sci. USA* **111**, 7898–7905 (2014).
6. Fink, M. *et al.* A neuronal two P domain K⁺ channel stimulated by arachidonic acid and polyunsaturated fatty acids. *EMBO J.* **17**, 3297–3308 (1998).
7. Noël, J. *et al.* The mechano-activated K⁺ channels TRAAK and TREK-1 control both warm and cold perception. *EMBO J.* **28**, 1308–1318 (2009).
8. Brohawn, S. G., Su, Z. & MacKinnon, R. Mechanosensitivity is mediated directly by the lipid membrane in TRAAK and TREK1 K⁺ channels. *Proc. Natl Acad. Sci. USA* **111**, 3614–3619 (2014).
9. Kung, C., Martinac, B. & Sukharev, S. I. Mechanosensitive channels in microbes. *Annu. Rev. Microbiol.* **64**, 313–329 (2010).
10. Sukharev, S. I., Blount, P., Martinac, B. & Kung, C. Mechanosensitive channels of *Escherichia coli*: the MscL gene, protein, and activities. *Annu. Rev. Physiol.* **59**, 633–657 (1997).
11. Perozo, E., Cortes, D. M., Somporpnisut, P., Kloda, A. & Martinac, B. Open channel structure of MscL and the gating mechanism of mechanosensitive channels. *Nature* **418**, 942–948 (2002).
12. Brohawn, S. G., del Mármol, J. & MacKinnon, R. Crystal structure of the human K2P TRAAK, a lipid- and mechano-sensitive K⁺ ion channel. *Science* **335**, 436–441 (2012).
13. Brohawn, S. G., Campbell, E. B. & MacKinnon, R. Domain-swapped chain connectivity and gated membrane access in a Fab-mediated crystal of the human TRAAK K⁺ channel. *Proc. Natl Acad. Sci. USA* **110**, 2129–2134 (2013).
14. Miller, A. N. & Long, S. B. Crystal structure of the human two-pore domain potassium channel K2P1. *Science* **335**, 432–436 (2012).
15. del Camino, D. & Yellen, G. Tight steric closure at the intracellular activation gate of a voltage-gated K⁺ channel. *Neuron* **32**, 649–656 (2001).
16. Jiang, Y. *et al.* The open pore conformation of potassium channels. *Nature* **417**, 523–526 (2002).
17. Harinath, S. & Sikdar, S. K. Trichloroethanol enhances the activity of recombinant human TREK-1 and TRAAK channels. *Neuropharmacology* **46**, 750–760 (2004).
18. Phillips, R., Ursell, T., Wiggins, P. & Sens, P. Emerging roles for lipids in shaping membrane-protein function. *Nature* **459**, 379–385 (2009).
19. Opsahl, L. R. & Webb, W. W. Lipid–glass adhesion in giga-sealed patch-clamped membranes. *Biophys. J.* **66**, 75–79 (1994).
20. Ursell, T., Kondev, J., Reeves, D., Wiggins, P. A. & Phillips, R. In *Mechanosensitive Ion Channels* (eds Kamkin, A. & Kiseleva, I.), **1**, 37–70 (Springer, 2008).
21. Enyedi, P. & Czirják, G. Molecular background of leak K⁺ currents: two-pore domain potassium channels. *Physiol. Rev.* **90**, 559–605 (2010).
22. Honoré, E., Patel, A. J., Chemin, J., Suchyna, T. & Sachs, F. Desensitization of mechano-gated K2P channels. *Proc. Natl Acad. Sci. USA* **103**, 6859–6864 (2006).
23. Maksaev, G., Milac, A., Anishkin, A., Guy, H. R. & Sukharev, S. I. Analyses of gating thermodynamics and effects of deletions in the mechanosensitive channel TREK-1: comparisons with structural models. *Channels* **5**, 26–34 (2011).
24. Chemin, J. *et al.* A phospholipid sensor controls mechanogating of the K⁺ channel TREK-1. *EMBO J.* **24**, 44–53 (2005).
25. Murbartián, J., Lei, Q., Sando, J. J. & Bayliss, D. A. Sequential phosphorylation mediates receptor- and kinase-induced inhibition of TREK-1 background potassium channels. *J. Biol. Chem.* **280**, 30175–30184 (2005).
26. Patel, A. J. *et al.* A mammalian two pore domain mechano-gated S-like K⁺ channel. *EMBO J.* **17**, 4283–4290 (1998).
27. Honoré, E., Maingret, F., Lazdunski, M. & Patel, A. J. An intracellular proton sensor commands lipid- and mechano-gating of the K⁺ channel TREK-1. *EMBO J.* **21**, 2968–2976 (2002).
28. Bagriantsev, S. N., Clark, K. A. & Minor, D. L. Metabolic and thermal stimuli control K2P2.1 (TREK-1) through modular sensory and gating domains. *EMBO J.* **31**, 3297–3308 (2012).
29. Aryal, P., Abd-Wahab, F., Bucci, G., Sansom, M. S. P. & Tucker, S. J. A hydrophobic barrier deep within the inner pore of the TWIK-1 K2P potassium channel. *Nature Commun.* **5**, 4377 (2014).

Acknowledgements We thank staff at APS beamlines 23-IDB/D, especially R. Sanishvili and S. Corcoran and at 24-IDC/E, especially I. Kourinov, D. Neau and K. Rajashankar for assistance at the synchrotron and members of the MacKinnon laboratory for discussions. S.G.B. is a Howard Hughes Medical Institute postdoctoral fellow of the Helen Hay Whitney Foundation and R.M. is an investigator of the Howard Hughes Medical Institute.

Author Contributions S.G.B. designed and performed the experiments. E.B.C. performed hybridoma cell culture. R.M. supervised the project. S.G.B. and R.M. analysed the data and wrote the manuscript.

Author Information The X-ray crystallographic coordinates and structure factors for the non-conductive conformation of TRAAK in K⁺ (4WFF) and TI⁺ (4WFH) and the conductive conformation of TRAAK in K⁺ (4WFE) and TI⁺ (4WFG) are available at the Protein Data Bank. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.M. (mackinn@rockefeller.edu).

METHODS

Protein expression and purification. Cloning of a codon-optimized variant of the *Homo sapiens* TRAAK gene (UniProt Q9NYG8-2) and heterologous expression in *Pichia pastoris* was previously described^{8,12,13}. The construct used in this study is C-terminally truncated by an additional ten amino acids compared to our previous reports and similarly incorporates two mutations to remove N-linked glycosylation sites (N104Q/N108Q) and is expressed as a C-terminal PreScission protease-cleavable EGFP–10×His fusion protein. TRAAK_{1–290(N104Q,N108Q)}–SNS–LEVLFG/GP–EGFP–H10 is referred to as TRAAK in the text for simplicity.

Frozen *Pichia* cells expressing TRAAK were disrupted by milling (Retsch model MM301) 5 times for 3 min at 25 Hz. All subsequent purification steps were carried out at 4 °C. For crystallization in K⁺, cell powder was added to lysis buffer (50 mM Tris pH 8.0, 150 mM KCl, 60 mM decyl- β -D-maltoside (DM, Affymetrix), 0.1 mg ml^{−1} DNase I, 1 μ g ml^{−1} pepstatin, 1 μ g ml^{−1} leupeptin, 1 μ g ml^{−1} aprotinin, 10 μ g ml^{−1} soy trypsin inhibitor, 1 mM benzamide, and 1 mM phenylmethylsulfonyl fluoride added immediately before use) at a ratio of 1 g cell pellet per 4 ml lysis buffer. Membranes were extracted for 3 h with stirring followed by centrifugation at 35,000g for 45 min. Cobalt resin (Clontech) was added to the supernatant (1 ml resin per 5 g cell pellet) and stirred gently for 3 h. Resin was collected on a column and serially washed and eluted in buffer (50 mM Tris pH 8.0, 150 mM KCl, 6 mM DM) with 10 mM, 30 mM, and 300 mM imidazole pH 8.0. EDTA pH 8.0 (1 mM final) and PreScission protease (~1:50 WT:WT) were added to the elution before incubation with gentle rocking overnight. Cleaved protein was concentrated (50 kDa MWCO) and applied to a Superdex 200 column (GE Healthcare) equilibrated in SEC buffer (20 mM Tris pH 8.0, 150 mM KCl, 1 mM EDTA, 4 mM *n*-decyl- β -D-maltopyranoside (DM)).

Fab was prepared from monoclonal antibody 13E9 against TRAAK as described¹³ and buffer exchanged on a HiTrap desalting column (GE Healthcare) into 20 mM Tris 150 mM KCl pH 8.0 for crystallization in KCl or 20 mM Tris 150 mM KNO₃ pH 8.0 for crystallization in TINO₃. TRAAK–Fab complexes were prepared by incubating purified channel concentrated (50 kDa MWCO) to ~10 mg ml^{−1} with purified Fab concentrated (10 kDa MWCO) to ~30 mg ml^{−1} at a 1:2.5 molar ratio in SEC buffer for 10 min at 4 °C. The TRAAK–Fab complex was separated from excess free Fab on a Superdex 200 column (GE Healthcare) equilibrated in SEC buffer. TRAAK–Fab complexes were concentrated (10 kDa MWCO) to 30 mg ml^{−1} for crystallization.

For crystallization in TI⁺, TRAAK–Fab complexes were purified identically, except that 150 mM KNO₃ replaced the 150 mM KCl in all buffers except for the final SEC buffer in which 150 mM TINO₃ replaced the 150 mM KCl.

Crystallization, data collection, and structure determination. Crystals were grown in drops of 0.25–0.35 μ l protein added to an equal volume of reservoir, in hanging drops over a 100 μ l reservoir at 4 °C. Reservoir for the non-conductive conformation in K⁺ was 50 mM Tris pH 8.8, 200 mM CaCl₂, 27–30% (vol/vol) PEG 400. Reservoir for the non-conductive conformation in TI⁺ was 50 mM TRIS pH 8.8, 200 mM Ca(NO₃)₂, 27–30% (vol/vol) PEG 400. Reservoir for the conductive conformation in K⁺ was 50 mM Tris pH 8.8, 200 mM CaCl₂, 27–30% (vol/vol) PEG 400, 4% (vol/vol) polypropylene glycol 400. Reservoir for the conductive conformation in TI⁺ was 50 mM Tris pH 8.8, 200 mM Ca(NO₃)₂, 27–30% (vol/vol) PEG 400, 52 mM trichloroethanol. Crystals grew to ~100 μ m × 100 μ m × 200 μ m in 4–6 weeks. For cryoprotection, an approximately equal volume of mother liquor supplemented to be 30% (vol/vol) PEG 400 was added to one side of the drop and crystals were moved through this solution with a cryoloop before being plunged into liquid nitrogen.

Data were collected at beamline 23-IDB or 24-IDC at the APS. Thallium-containing crystals were collected at the energy giving a maximum imaginary scattering component f'' determined by a fluorescent scan. Data were processed with HKL2000 (ref. 30) and structures were solved by molecular replacement with Phaser³¹ using the previously solved TRAAK–Fab structure¹³ with ligands, TM2–TM3 and TM4 regions removed as the search model. Consistent with recent analysis³², incorporation of weak data at high resolution (to CC_{1/2} = 35–60%) improved map quality. The structures were modelled in Coot³³ and refined in Refmac³⁴. Molprobity³⁵ was used to assess model geometry. Twofold local NCS and jelly body restraints were used throughout refinement and three TLS groups per protein chain were incorporated during the final rounds of refinement. 1,350 of 1,454 total amino acids (93%) are modelled in the structures: the termini (amino acids 1–26 and 285–290 plus a nine amino acid linker) and a loop connecting the helical cap to pore helix 1 (amino acids A104–A109 and B106–B109) in each TRAAK protomer and a loop in each Fab heavy chain (amino acids 130–135) were not visible in the electron density. In the non-conductive structure in TI⁺, residues 275–284 from TRAAK subunit B were additionally not modelled.

The positions of seven TI⁺ ions in the non-conductive conformation (two bound to the extracellular surface of TRAAK, one above the cytoplasmic mouth of the filter, and four in the selectivity filter) and eight TI⁺ ions in the conductive conformation (with an additional TI⁺ in the cavity) were identified in model phased anomalous difference Fourier maps as the highest peaks above background. PhaserEP³⁶ automatically

found the same sites. The presence of an acyl chain in the channel cavity was evaluated based on the shape of the electron density (in omit maps) and statistics after crystallographic refinement. The mean B-factor and real space correlation coefficient of the acyl chain (101 Å²/0.76 in the K⁺ structure and 132 Å²/0.74 in the TI⁺ structure) are comparable to the values of nearby residues in TM2 and TM4 that project into the cavity (amino acids 154, 155, 158, 161, 269, 272, 273, and 276 have an average B-factor / real space correlation coefficient of 125 Å²/0.78 in the K⁺ structure and 142 Å²/0.83 in the TI⁺ structure).

Phases for omit maps were calculated from a ligand-free model with every atom randomly shifted by a maximum of 0.5 Å before 20 cycles of refinement in Refmac to remove model bias. Omit density in these maps was essentially indistinguishable from maps generated with naive models that never had ligands modelled or simulated annealing omit maps calculated with Phenix.

Structures were compared after alignment of pore helices and selectivity filters (119–133 and 228–242 from chain A and B) using LSQKAB in CCP4. Cross-sectional area calculations were performed with CHARMM^{37,38}. The area calculation used a surface calculated with 1.4 Å added to Van der Waals radii of protein atoms and a probe radius of 3.5 Å to approximate the lipid-accessible surface area. A water cylinder with a 6 Å radius was used to fill the cavity of the TRAAK channels to exclude its contribution from the calculation. All area calculations and structures in Figs 3 and 4 were made with symmetric TRAAK channels (that is, both subunits had the same conformation) generated by rotation of a copy of one subunit onto the other. Internal pore diameters were calculated with the program Hole³⁹. Structure figures were made with Pymol (Version 1.7.2 Schrödinger, LLC) or the UCSF Chimera⁴⁰ package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).

Electrophysiology. For cellular recording, CHO-K1 cells (ATCC) (which do not express endogenous mechanosensitive channels⁶) were cultured in DMEM-F12 (Gibco) with 10% FBS, 0.2 mM L-alanyl-L-glutamine (Glutamax, Gibco), 100 U ml^{−1} penicillin, and 100 μ g ml^{−1} streptomycin. Cells were plated in 35 mm polystyrene dishes (BD Biosciences) or on polyester Thermanox coverslips (Thermo Scientific) ~24 h before transfection with FugeneHD (Promega) following the manufacturer's protocol. After 48 h, dishes or coverslips were transferred to a recording chamber and cells were washed with bath solution. DNA encoding amino acids 1–290 of *H. sapiens* TRAAK was cloned into a pIRES2–AcGFP vector (Clontech) to generate untagged TRAAK_{1–290} used in experiments shown in Fig. 1a. DNA encoding amino acids 1–300 of *H. sapiens* TRAAK was cloned into the EcoRI/XhoI sites of a modified pCEH vector to generate a C-terminally EGFP tagged construct (TRAAK_{1–300}–SNSAVDAGLVPRGSA–EGFP–H10). This construct and mutants generated with inverse PCR were used in experiments presented in Fig. 2 and Extended Data Figs 3 and 4.

Reconstitution of TRAAK_{1–290(N104Q,N108Q)}–SNS–LEVLFG/GP–EGFP–H10 and recording from proteoliposome blisters was performed as described for TRAAK_{1–300(N104Q,N108Q)}–SNS–LEVLFG/GP–EGFP–H10 (ref. 8). Azolectin (egg L- α -phosphatidylcholine, 840051) and DPhPC (1,2-diphtanoyl-*sn*-glycero-3-phosphocholine, 850356) lipids were from Avanti Polar Lipids.

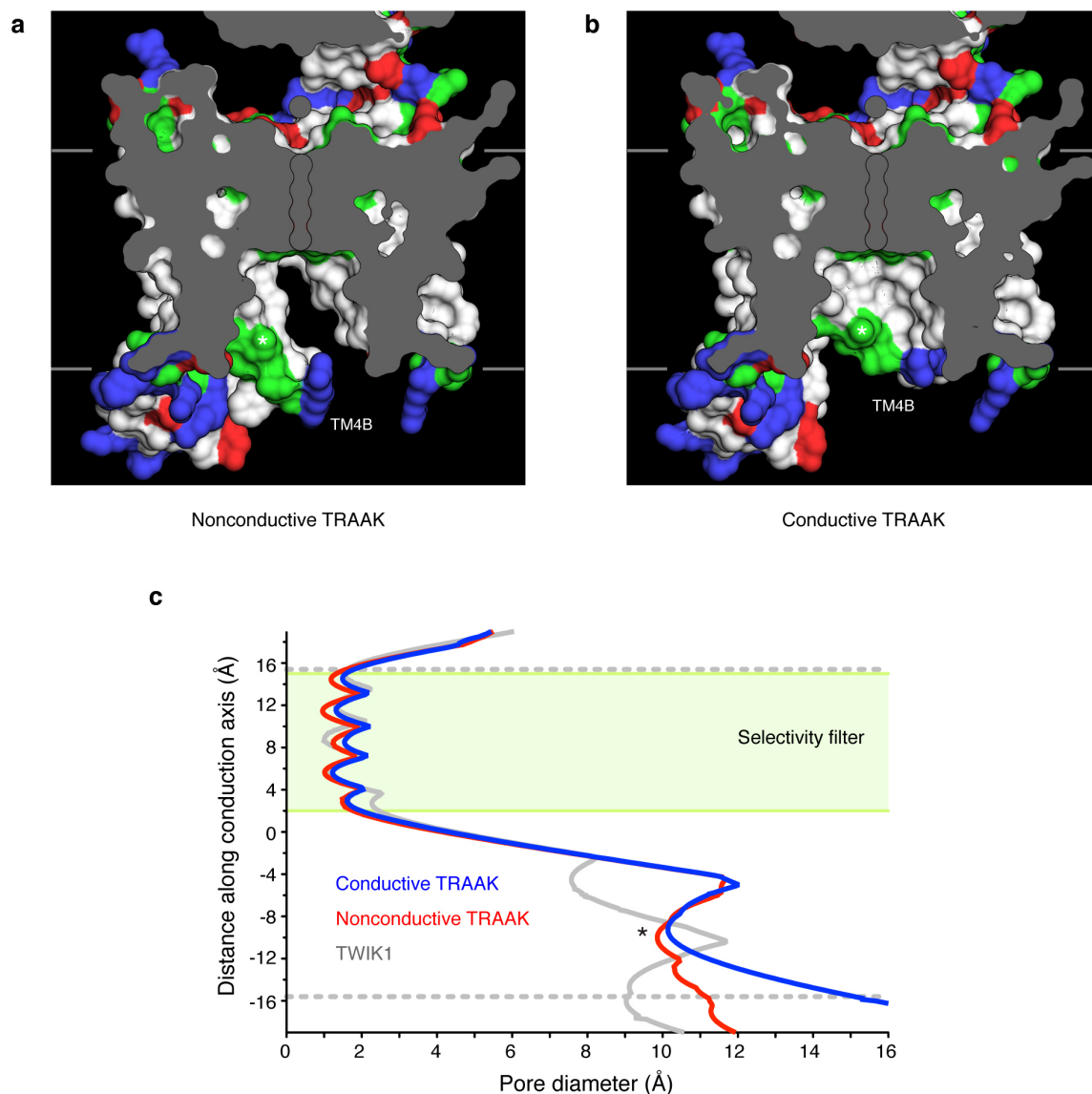
Pipettes were pulled from borosilicate glass (Sutter Instruments BF150-86-10) to 3–4 M Ω resistance. Recordings were made with an Axopatch 200B amplifier (Molecular Devices), filtered at 1 kHz and digitized at 10–100 kHz (Digidata 1440A, Molecular Devices). All recordings were performed at room temperature. Pressure application through the patch pipette was accomplished with a high-speed pressure clamp (HSPC, ALA Scientific) controlled through the Clampex software. Pressure application velocity was set to the maximum rate of 8.3 mm Hg msec^{−1}.

All recordings were performed in a tenfold concentration gradient of K⁺. For cellular recordings, pipette solution was 10 mM Hepes, 150 mM KCl, 3 mM MgCl₂, 5 mM EGTA, pH 7.2 (adjusted with KOH) and bath solution was 10 mM Hepes, 15 mM KCl, 135 mM NaCl, 3 mM MgCl₂, 1 mM CaCl₂, pH 7.3 (adjusted with NaOH). For experiments in Fig. 2 and Extended Data Fig. 3, the reducing bath solution additionally contained 10 mM DTT added from a freshly thawed 1 M stock of DTT in water. Reducing solution was used within one hour of preparation. For recordings from proteoliposomes, pipette solution was 5 mM Hepes, 180 mM NaCl, 20 mM KCl, pH 7.2 (adjusted with NaOH) and bath solution was 5 mM Hepes, 200 mM KCl, 40 mM MgCl₂, pH 7.2 (adjusted with KOH). Perfusion was accomplished with a nitrogen pressurized micro-perfusion system (VC³-8×P, ALA Scientific). All recordings in the manuscript are presented such that positive currents indicate K⁺ flux from the high [K⁺] to low [K⁺] side.

30. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution from diffraction images to an initial model in minutes. *Acta Crystallogr. D* **62**, 859–866 (2006).

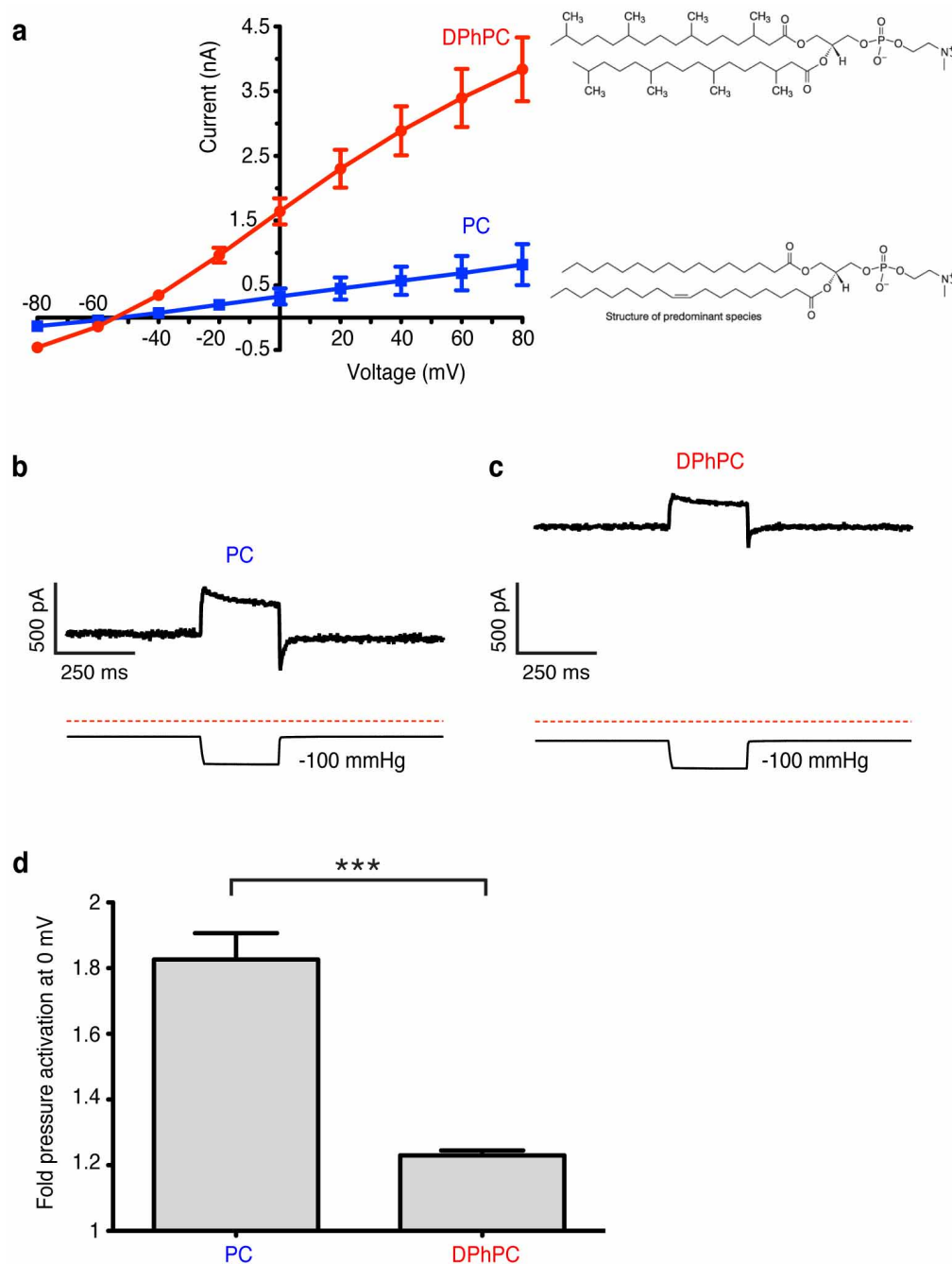
31. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

32. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
33. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
34. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
35. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst.* **D66**, 12–21 (2010).
36. Read, R. J. & McCoy, A. J. Using SAD data in Phaser. *Acta Crystallogr. D* **67**, 338–344 (2011).
37. Brooks, B. R. *et al.* CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
38. Jo, S., Kim, T. & Im, W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS ONE* **2**, e880 (2007).
39. Smart, O. S., Goodfellow, J. M. & Wallace, B. A. The pore dimensions of gramicidin A. *Biophys. J.* **65**, 2455–2460 (1993).
40. Pettersen, E. F., Goddard, T. D. & Huang, C. C. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Computational Chem.* **25**, 1605–1612 (2004).



Extended Data Figure 1 | The central cavity in conductive and non-conductive TRAAK conformations. **a, b,** View from the membrane plane of the TRAAK central cavity in the non-conductive (**a**) and conductive (**b**) conformations. The exposed surface of hydrophobic amino acids are colored white, arginine and lysine are blue, glutamate and aspartate are red, and polar residues are green. The positions of K^+ ions in the filter are outlined and residue T277 in TM4 is indicated with an asterisk. **c,** Diameter of the ion

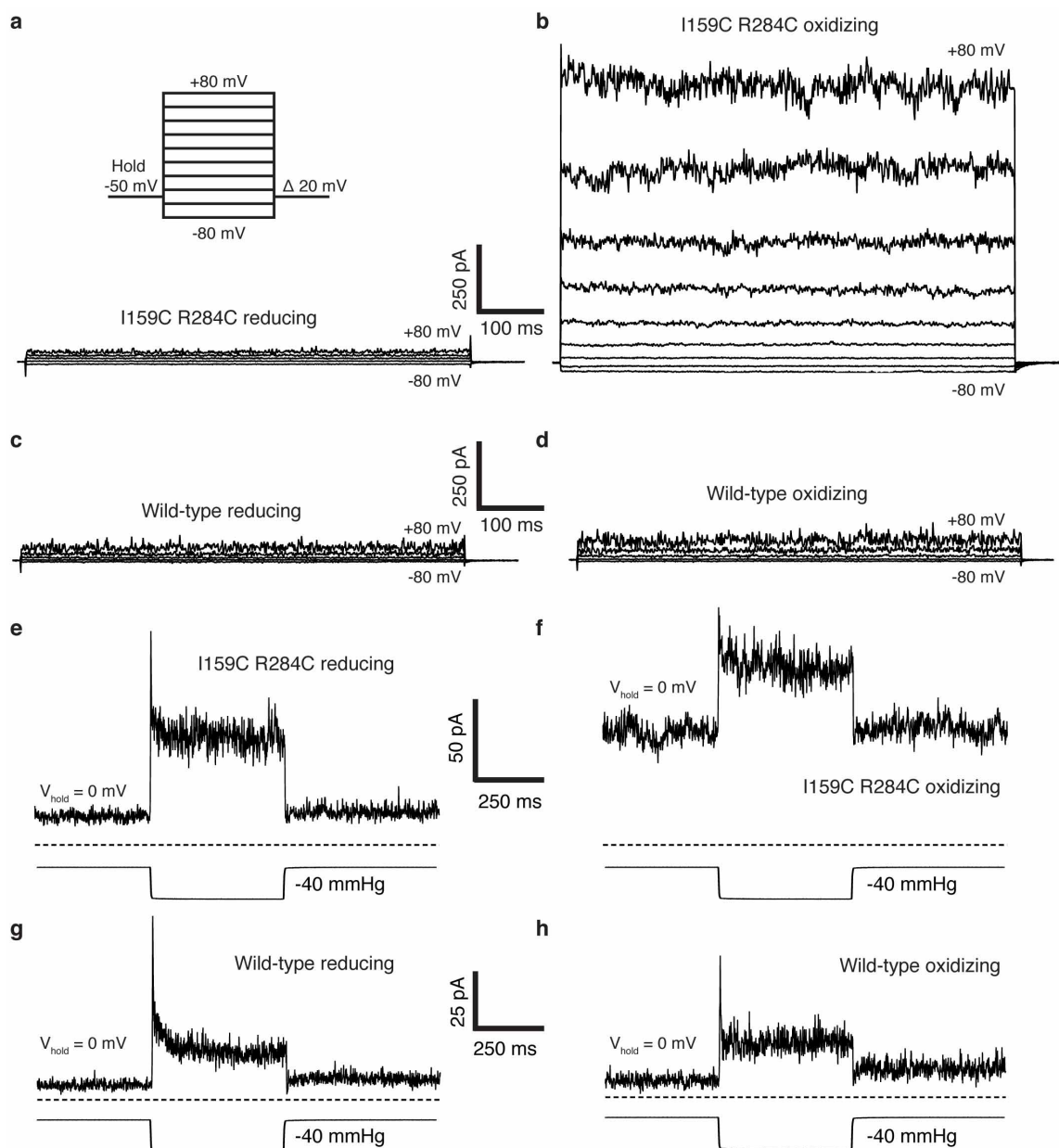
conduction pathway as a function of distance through the membrane for non-conductive TRAAK (red), conductive TRAAK (blue) and TWIK1 (grey, PDB: 3UKM). The green box indicates the position of the selectivity filter and dashed grey lines are the approximate boundaries of the lipid membrane. The ~ 10 Å diameter constriction formed partially by T277 is indicated with an asterisk. The pore diameter is larger in TRAAK than in TWIK1 and expands below T277 in the conductive conformation.



Extended Data Figure 2 | Reconstituted TRAAK activity in different lipids.

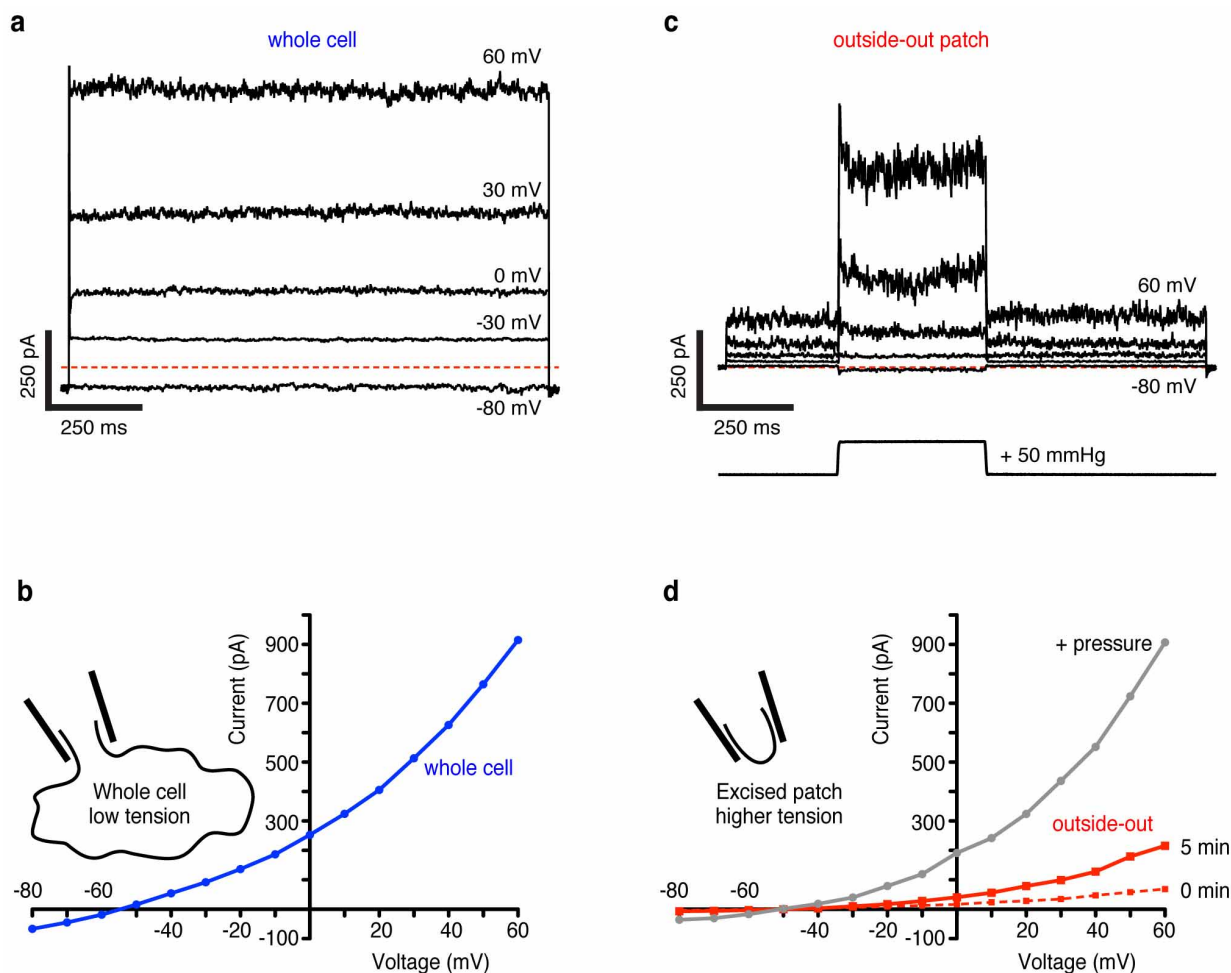
a, Current recorded from TRAAK proteoliposome patches as a function of holding voltage (mean \pm s.e.m., $n = 9$ patches each from three separate experiments). Current through TRAAK reconstituted in phosphatidylcholine lipids with branched acyl chains (1,2-diphytanoyl-*sn*-glycero-3-phosphocholine, DPhPC) was significantly higher than in non-branched acyl

chains (egg 1- α -phosphatidylcholine, PC) at each voltage measured (5.0-fold higher at 0 mV, $P < 0.0001$, Student's t -test). **b**, **c**, Representative recording of pressure (lower trace) activation of TRAAK current (upper trace) in PC (**b**) or DPhPC (**c**) lipids. **d**, Quantification of pressure activation of TRAAK in PC and DPhPC (mean fold pressure activation at 0 mV \pm s.e.m., $n = 9$ patches each from three separate experiments, *** $P < 0.0001$, Student's t -test).



Extended Data Figure 3 | Representative electrophysiological recordings from wild-type TRAAK and TRAAK(I159C, R284C). In these experiments and those in Fig. 2, inside-out patches from cells expressing wild-type or mutant TRAAK channels were excised and perfused with reducing bath solution (with 10 mM DTT). After stabilization of the patch (TRAAK channels exhibit a gradual run-up of current following excision to an equilibrium value, for example, Extended Data Fig. 4), the perfusion solution was switched to oxidizing bath solution (no DTT). **a, b**, Representative voltage family from a

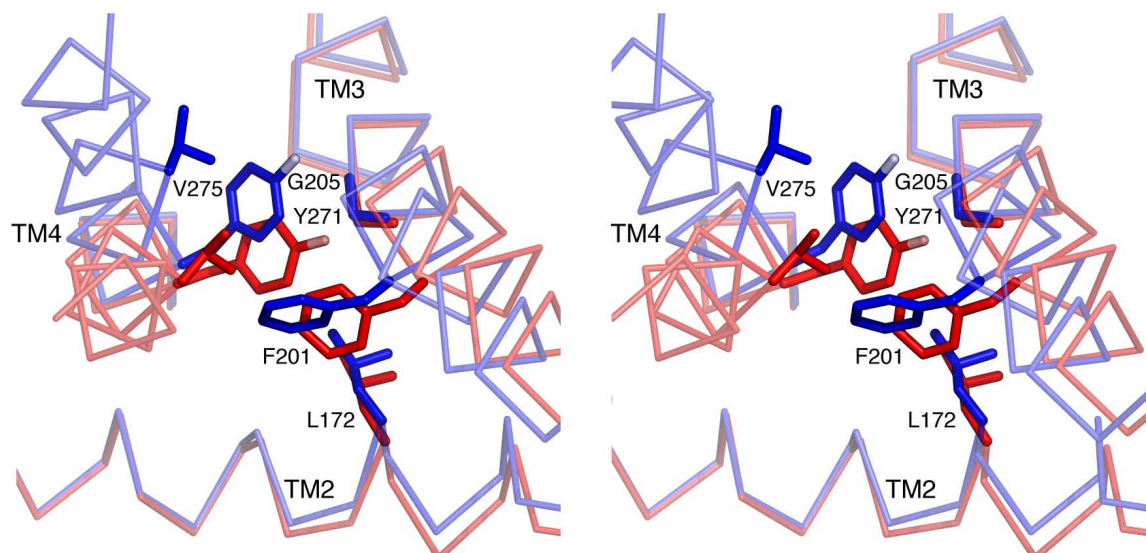
TRAAK(I159C, R284C) patch during perfusion of reducing (**a**) and oxidizing (**b**) solution. The voltage family protocol is illustrated. **c, d**, Same as **a, b**, but from a wild-type TRAAK patch. **e, f**, Current response (upper) to pressure application (lower) at 0 mV from the same TRAAK(I159C, R284C) patch during perfusion of reducing (**e**) or oxidizing (**f**) bath solution. **g, h**, Same as **e, f**, but from a wild-type TRAAK patch. Scale is shown between each pair of recordings in reducing and oxidizing bath solutions.



Extended Data Figure 4 | Basal activity and tension activation of TRAAK.

a, Whole cell current from a TRAAK-expressing cell during a voltage step protocol in a tenfold gradient of $[K^+]$ ($E_{K^+} = -59$ mV, holding voltage = -80 mV, $\Delta V = 10$ mV, indicated steps shown). Red dashed line indicates zero current level. **b**, Current–voltage relationship from experiment in **a**. **c**, Currents (upper traces) recorded from an outside-out patch excised from the same cell as in **a**, **b**. The voltage protocol in **a** was used with an additional pressure step (lower trace) during each voltage step. **d**, Current–voltage relationship from data in **b** (mean current 5 min after patch excision before pressure (red)

and peak current during pressure step (grey)) and a recording immediately after pulling the patch (red dashes). The excised patch contains $< 1\%$ of the whole cell membrane area, but gives $\sim 25\%$ of the whole cell current before and similar current during a pressure step. This is explained by very low basal activity of TRAAK with near-zero membrane tension (whole cell) and channel activation by increasing membrane tension over a broad range (intermediate tension in an excised patch to high tension in a pressurized patch).



Extended Data Figure 5 | Detailed view of TM2-TM3 rotation in TRAAK. Stereo view from the cytoplasm of an overlay of non-conductive (red) and conductive TM2-TM3 rotated (blue) conformations. Amino acids that sterically prevent TM2-TM3 rotation when TM4 is down are shown as sticks. TM2-TM3 rotates about hinges at positions G169 and G205. This rotation can

only occur if TM4 is up because amino acids L172, F201 and G205 on TM2-TM3 shift (0.75–2.1 Å) to a position that would sterically clash with amino acids Y271 and V275 on TM4 in a down conformation. Translation of Y271 and V275 3.1–4.1 Å in TM4 up conformations creates space for the TM2-TM3 rotation.

Extended Data Table 1 | Data collection and refinement statistics

	Conductive K ⁺ bound PDB: 4WFE	Conductive TI ⁺ bound PDB: 4WFG	Nonconductive K ⁺ bound PDB: 4WFF	Nonconductive TI ⁺ bound PDB: 4WFH
Data collection				
Space group	P2 ₁	P2 ₁	P2 ₁	P2 ₁
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	80.8, 138.5, 96.8	80.8, 138.8, 96.3	80.7, 138.9, 96.5	81.0, 138.8, 96.8
α , γ , β (°)	90, 95.1	90, 94.6	90, 95.1	90, 94.6
Wavelength (Å)	1.0332	0.9781	1.0332	0.9781
Resolution (Å)	50-2.5 (2.54-2.5)*	50-3.0 (3.05-3)	50-2.5 (2.54-2.5)	50-3.0 (3.05-3)
<i>R</i> _{sym} (%)	6.3 (>100)	9.7 (>100)	9.3 (>100)	15.3 (>100)
<i>R</i> _{plim} (%)	4.3 (97.6)	5.8 (98.6)	4.0 (94.2)	7.0 (97.9)
<i>CC</i> _{1/2} (%)	(35.1)	(58.6)	(44.5)	(44.2)
<i>I</i> / σ <i>I</i>	22.6 (0.9)	14.7 (0.9)	34.8 (1.0)	11.3 (0.6)
Completeness (%)	99.8 (99.8)	99.6 (99.7)	99.9 (100)	100 (100)
Redundancy	3.8 (3.7)	5.6 (5.7)	10.1 (10.3)	10.5 (10.7)
Refinement				
Resolution (Å)	48.2 - 2.5	48.0 - 3.0	48.1 - 2.5	48.2 - 3.0
No. reflections†	65770	40099	68997	39746
<i>R</i> _{work} / <i>R</i> _{free}	19.9 / 23.4	20.2 / 23.6	21.3 / 24.5	20.3 / 23.9
No. atoms				
Total	10598	10599	10585	10506
K ⁺ / TI ⁺	6	8	5	7
Decane	0	0	10	10
B-factors				
All atoms	90	109	98	112
K ⁺ / TI ⁺	89	94	137	170
Decane	n/a	n/a	101	131
R.m.s deviations				
Bond lengths (Å)	0.008	0.006	0.008	0.006
Bond angles (°)	1.058	0.898	1.002	0.908
Ramachandran				
favored (%)	96.69	97.37	97.14	96.51
allowed (%)	3.16	2.48	2.71	3.26
disallowed (%)	0.15	0.15	0.15	0.23

Data from 1, 1, 3 and 2 crystals were merged for the conductive K⁺, conductive TI⁺, non-conductive K⁺ and non-conductive TI⁺ structures, respectively.

* Values for the highest resolution shell are shown in parenthesis.

† 5% of these reflections were used to calculate *R*_{free}.

Extended Data Table 2 | Pairwise root mean square deviation (Å) between conformationally static regions between TRAAK structures

	Conductive K ⁺ bound PDB: 4WFE	Nonconductive K ⁺ bound PDB: 4WFF	Conductive TI ⁺ bound PDB: 4WFG
Nonconductive K ⁺ bound PDB: 4WFF	0.217	-	-
Conductive TI ⁺ bound PDB: 4WFG	0.186	0.227	-
Nonconductive TI ⁺ bound PDB: 4WFH	0.221	0.137	0.193

Conformationally static regions of the TRAAK channel in these structures include the entire A chain and residues 1–155 of chain B.

Extended Data Table 3 | Summary of TRAAK channel conformations in crystal structures

State as a homodimer	PDB ID	Chain ID	TM4 inner half conformation	TM2-TM3 conformation	TM4 outer half conformation
Nonconductive	4I9W	B	Down	Out	Up
	4WFF	B	Down	Out	Up
	4WFH	B	Down	Out	Up
	3UM7	B	Down	Out	Up
	3UM7	A	Down	Out	Up
Conductive	4WFE	A	Up	In	Displaced
	4WFF	A	Up	In	Displaced
	4WFG	A	Up	In	Displaced
	4WFH	A	Up	In	Displaced
	4I9W	A	Up	In	Displaced
	4WFE	B	Up	Out	Up
	4WFG	B	Up	Out	Displaced

Structures shown in Fig. 3 are highlighted with the colour used in the figure.

TOOLBOX THE AUTOMATED LAB

Start-up firms say robotics and software that autonomously record every detail of an experiment can transform the efficiency and reliability of research.

ILLUSTRATION BY THE PROJECT TWINS



BY ERIKA CHECK HAYDEN

Max Hodak has spent much of his academic career fixing the ways that scientists collect data. As a biomedical engineering student at Duke University in Durham, North Carolina, it frustrated him that his laboratory recorded its experiments in paper notebooks, leaving researchers to scour through the pages to find relevant data. So in 2008, he indexed all the notebook data on a computer

and wrote a program to allow users to query it. “People were saying, ‘Why are you wasting your time? That’s not going to lead to publication,’” he recalls. But a year-and-a-half later, he returned to the lab from a stint in Silicon Valley to find that many of those earlier sceptics were now using his system. To Hodak, it was a sign that he should pursue his quest for efficiency in the lab. “I was always more interested in finding ways to do analysis more efficiently than in doing the actual analysis,” he says.

Today, a warehouse in California is the living embodiment of Hodak’s dream to build an automated lab that conducts experiments and records the results, or what he calls a “biology data centre”. His company, Transcriptic, founded in 2012, is the first of a crop of start-ups of this ilk, all with a similar claim: that advances in software and robotics will help to free researchers from manual drudgery, make their data easier to store and query, and ultimately lead to cheaper, more ►

► efficient and more reproducible science.

Transcriptic and another California firm called Emerald Therapeutics are pinning their hopes on offering scientists control of a wet lab by remote computer. Many big biology labs already have automated machines to sequence or copy pieces of DNA. But these companies want to bring automation to other routine experiments, such as moving and separating proteins or fragments of DNA through a gel. They offer the capabilities to labs big and small.

In Transcriptic's approach, customers first program an experiment using the company's application programming interface (API), which translates each step of an experimental protocol into machine-readable code. The customer's orders — and any physical samples — arrive at Transcriptic's warehouse in Menlo Park, and the experiment is carried out in a Plexiglas-enclosed station that contains a benchtop full of instruments guided by a computer, which receives the work order and runs the assemblage of machines. A robot on a gantry runs the length of the station, transferring plates from machine to machine — apparatus for the polymerase chain reaction (which amplifies DNA), plate readers, liquid handlers, a freezer and an incubator — to carry out the experiment. Users receive data from each step of the experiment in real time.

Customers say that this frees them up to spend their time on science, rather than on grunt work. For example, synthetic biologist Justin Siegel heads a lab at the University of California, Davis, that designs, builds and tests new enzymes. Transcriptic now does the lab's molecular biology, liberating his students from the one-third of their time that used to be spent on copying and mutating DNA fragments (cloning and mutagenesis, respectively). Siegel credits Transcriptic with building a biosensor that detects the chemical profile of olive oil, for which his students won the grand prize at the 2014 International Genetically Engineered Machines competition on 3 November. "It's made us more efficient and a little bolder," he says. "Instead of making just ten designs, they want to try a couple extra. They'll go a little bit farther out on a limb because all of a sudden they don't have to physically build the stuff."

EXPERIMENTS TO ORDER

Emerald is testing what it calls the Emerald Cloud Laboratory, which the company says will be a one-stop online shop through which customers order experiments, analyse data and collaborate with others. Starting early in 2015, beta users will be able to order from a list of 40 common lab protocols, such as western blots for protein analysis, or high-performance liquid chromatography for separating components of a mixture. When an order comes in, a human operator will set up the experiment on one of the company's automated workstations

at its lab in Menlo Park. Operators transfer sample plates from machine to machine to carry out the steps of the experiment, and the customer gets data back through the Emerald Cloud Laboratory. There, customers can analyse the results using text functions encoded in the Wolfram programming language. Users can review everything, from the controls and machine settings of the instruments used, to results from past experiments using the same reagents — even experiments from others who have granted permission for their data to be openly accessible. "Everything we do is built by scientists for scientists," says Emerald co-founder Brian Frezza.

In one sense, these models are similar to conventional contract-research organizations, but the automated systems and data collection offer scientists much finer detail and control over experimental design. One might think that the complicated equipment makes experiments

more expensive to do. But Hodak says that Transcriptic offers protocols such as cloning and mutagenesis at about the same cost at which they could be run in an academic lab — or for less — and at about half the price offered by conventional research outsourcing firms, he says. That is partly because its stations can run without a human operator, and also because the firm makes a lot of its own hardware. When Hodak first went looking for an automated freezer, the cheapest he could find was US\$400,000. So he hired mechanical engineers to build one for \$40,000, and used the same basic design for automated incubators and refrigerators.

Although these labs are powerful research aids, remote users are necessarily limited to a standard set of experiments and instruments, notes Roger Chen, an associate at the investment firm O'Reilly AlphaTech Ventures in San Francisco, California. "I have a hard time believing that a centralized automated lab will give you the freedom and flexibility to experiment with all the parameters you need to do some innovation."

Chen has therefore invested in Riffyn, another fledgling start-up, in Oakland, California. It wants to adapt individual research labs so that scientists can use automated data collection for their own custom experiments, albeit without robotic control. The firm (which will not launch until late 2015 at the earliest) is building a cloud-computing

► **NATURE.COM**
For more on scientific software, apps and online tools, visit:
nature.com/toolbox

software platform integrated with devices that stream data from lab equipment. According to co-founder Timothy Gardner, the software

will allow users to design workflows, analyse experimental data held on remote servers and thus change parameters (such as the temperatures or pressures at which an instrument operates) in response to the analysis of an experiment's performance. "We're trying to solve the problem of how you bring McDonald's-like efficiency to scientists without shackling them to McDonald's-like recipes," he says.

COMMUNICATION IS KEY

Current scientific instruments take instructions and record data in a variety of formats. Harmonizing their software will be an even harder technical challenge than building robotically controlled equipment, warns Frezza. Gardner acknowledges this, but says that momentum for a common, open set of standards and software is building. Meanwhile, he hopes that Riffyn's software will enable lab devices to talk to each other — and to scientists — more smoothly. The concept would not have been possible even a few years ago; only last year did computer giant Apple, for instance, release iBeacon, a system that enables nearby compatible devices to communicate with each other.

Gardner's inspiration comes from his time running research operations at synthetic-biology pioneer Amyris Biotechnologies, in nearby Emeryville, which genetically engineers yeast to produce biofuels and specialty chemicals. In the early days it struggled to scale up basic processes to industrial quantities, because random variation — noise — in each step led to irreproducible results. The company started to analyse data from each step of the process to spot the weak points, and made dramatic improvements. Gardner says that research labs need that same kind of reproducibility. "We need to bring the pursuit of precision reliability to the academic world," agrees Douglas Crawford, associate director of the California Institute for Quantitative Biosciences (QB3), and one of Gardner's strongest supporters.

Ultimately, Gardner and other backers of automated labs say that the immediate payoff of their work might be to promote a general movement to boost the overall quality of research. Tools that make it easy for scientists to monitor and record every aspect of their experiment, they say, might help to deal with what some argue is a 'reproducibility crisis' in research — the sense that many experiments are too sloppily done, or that methods and data are recorded too imprecisely, for others to easily reproduce findings.

Crawford thinks that changing this will require a cultural shift as much as a technical one. Gardner agrees, but hopes that companies such as his will remove one of the roadblocks: "I don't think you can get the cultural and educational changes to stick if you don't have tools that make it easy," he says. ■

CAREERS

ENROLMENT United States is drawing more international graduate students **p.135**

GRANTS UK biomedical funder targets young and daring research **p.135**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



INDUSTRY SKILLS

Boosting business

Programmes prepare biomedical trainees for the management side of science.

BY KENDALL POWELL

Dana Watt represents a new generation of cell biologists. Even when she started her PhD in autumn 2010, she knew she was not heading for an academic career. Graduate students today, she says, are quite aware that they will not all become tenured professors. “We either have to ready ourselves for the long postdoctoral haul, or we can explore options outside of academia.” So she jumped when her university posted an

announcement about an off-campus short course on biotechnology business management during her fourth year of doctoral studies. She soaked up the new language and approach to scientific discovery.

A broad array of jobs outside academia requires business know-how — and there are many more of these jobs than there are conventional bench-research positions in biotechnology and pharmaceutical firms. A case in point: only a dozen of the 340 jobs posted by the biotechnology company Genzyme in

South San Francisco, California, in 2011 were strictly research-and-development positions, according to an outside analysis. The other postings were mainly for positions in regulatory affairs, marketing, quality assurance or management — all of which require business skills beyond what most doctoral students acquire in their degrees. The abundance is tantalizing, but it is difficult and stressful to discern or develop a career path to these jobs.

“The PhD comes off as very one-size-fits-all,” says Watt, now in the fifth year of her PhD at Washington University in St. Louis, Missouri, and planning to pursue a career in biotechnology consulting. “We need to rethink the way we train PhDs. I had to look to extra-curricular sources for training because ‘not academic’ is not a real career path.”

To help students in cell biology and other biomedical disciplines to navigate these non-academic career paths, various short courses and boot-camp-style workshops offer auxiliary business training. They showcase the vast range of opportunities in the biotechnology and pharmaceutical arenas. And — in sharp contrast to master’s programmes in business administration (MBAs), which typically require two years of costly study — these courses are short and usually inexpensive. For those who want more intense training, at least one US campus has started to offer business curricula tailored to people with life-sciences doctorates. These programmes give people with PhDs — who already have valuable analytical skills — the vocabulary and tools to bridge to the business world.

Early-career researchers will find that many of these offerings are aimed at their demographic. They can learn about them through their institution’s graduate division or postdoctoral office, or at conferences, where scientific societies often host workshops. The best programmes include lectures by and networking opportunities with industry leaders, and direct connections to companies, including internship offerings.

SHORT COURSES

Judging by its popularity, a course coordinated by the American Society for Cell Biology (ASCB) in Bethesda, Maryland, seems to have found the right format to give business training to biomedical scientists. Responding to what it saw as a business skills gap, last June the ASCB offered a 12-day course for graduate students and postdocs called Managing Science in the Biotech Industry. More than 500 people applied for the 40 all-expenses-paid spots at the Keck ►

► Graduate Institute in Claremont, California.

With strong financial support from the lab-supply company EMD Millipore, headquartered in Billerica, Massachusetts, the course featured Keck faculty members and industry leaders speaking on topics such as finance, intellectual property and project management.

The course's biotech business case studies fascinated participant Carlos Solórzano, a neuroscience postdoc at the University of California, San Francisco. "What kind of thought process does the CEO go through?" he says.

The course participants conducted a market analysis, choosing real biomarkers that could be developed into a cancer-diagnostic test. Using scientific, clinical and financial analyses, participants built a case for why a hypothetical company should invest in such a test.

Watt also attended, and was enthused after hearing a consultant speak about troubleshooting specific problems for companies. "The novelty and fast-paced aspects really fit my personality," she says. "Most importantly, I'm excited about a particular career path and how to get there." Solórzano came away from the course with a new appreciation of how to foster connections between a company and the outside world. The ASCB will offer the course again on 12–24 July 2015; applications are open from 4 December until 27 February.

For biomedical scientists who decide to commit to a business career, Keck also offers deeper training options. Founded in 1997, it is one of the few institutions worldwide that offer formal degree programmes in bioscience-industry skills. In the past five years it has begun to offer programmes that are tailored specifically to doctoral and postdoctoral students, such as a four-course certificate in bioscience management.

It also has a full-time professional master's programme for people with PhDs or medical doctorates (MDs). The US\$25,700 course takes nine months, with the option to do an internship afterwards. "We see this real valley — a skills gap — between basic research skills and skills needed for industry," says Steven Casper, dean of Keck's School of Applied Life Sciences. "To translate the scientific power of these scientists into useful skills in industry doesn't take a lot: four or five courses, working on a business project and doing an industry internship."

Keck's success — the full-time programme has a placement rate in salaried jobs of around 90% — lies mostly with its connections to biotech and pharmaceutical industry giants such as Pfizer, Genentech and Regeneron. Students spend the entire nine months working in teams on the business side of real company projects such as identifying a new application for a therapeutic compound.

Although the time and price tag might not appeal to PhD holders, Casper says that graduates start in positions with a median salary of \$80,000–90,000 and quickly advance. "They are super useful to these companies because

HELPING HANDS

Scratching the start-up itch

A well-worn path into the business world for researchers is to start companies in order to develop their own discoveries. When Lars Forsberg, a junior faculty member at Uppsala University in Sweden, found that loss of the Y chromosome in older men's blood cells was associated with an increased risk of cancer, he realized that he might have the basis for a business. He is relying heavily on two programmes in Sweden designed to help academic scientists to make the leap to industry: the Uppsala Innovation Centre (UIC) and Mentor4Research.

Forsberg's start-up cancer-diagnostics company, CRAY Innovation, sprang from a conversation that he had at the UIC, a business incubator closely associated with the university. The UIC patent adviser explained that if he did not patent the technology, no company would invest in it to develop it into a diagnostics kit. In Sweden, researchers — not universities — own the intellectual-property rights to their work.

"If I want this discovery to become something for the good of mankind, there's no other option than to do the business," says Forsberg. It helped that the UIC office was located in the building next to his lab.

The UIC provides the extra business

training that academic researchers need to commercialize their research. Forsberg participated in the centre's free three-month Business Lab course with lectures on business development, sales, financing and communication. At the end of the programme, participants write a business plan and pitch it to an investor panel.

Forsberg encourages other young researchers to seek out business incubators. He says that he has gained "the feeling of confidence when someone believes in you". And, just as importantly, these incubators often offer legal and financial aid to start-up companies.

Forsberg also participated in the Mentor4Research programme run by the Royal Swedish Academy of Engineering Sciences, which connects postgraduate researcher-entrepreneurs to experienced industry mentors.

"There is definitely a very big need for these kinds of programmes," says Anders Nordström, senior adviser for business development at the UIC and a regional manager of Mentor4Research. "Researchers don't have any idea about how to run the company or meet milestones or explain ideas to investors." **K.P.**

they can talk to any scientist or MD there, and then talk to the accountant."

Even though the Keck master's programme requires less time and financial investment than a conventional MBA, it is still a hefty outlay. Instead, Many biomedical trainees might just need exposure to the industry and the career options. Workshops let them sample the offerings.

HONING SKILLS

SciPhD, a career-development firm in Rockville, Maryland, offers a variety of one- and two-day programmes to universities, scientific societies and conferences across the United States and in Russia, Malaysia and Saudi Arabia. The cost per participant varies between \$60 and \$100, and some organizations refund the fees if students complete the course. SciPhD also offers a 40-hour, 2-month course through the New York Academy of Sciences for \$600–800 per participant.

"The training that scientists get as PhD students and postdocs is so disciplined, insightful, innovative and strategic — they are truly leaders" in a business sense already, says Randy Ribaud, co-founder of SciPhD. "They just don't know it."

His own career evolved from a lab-head position at the US National Cancer Institute in

Bethesda, Maryland, to a modelling-software company to a project manager at the health-care diagnostics firm Celera, where he picked up strategic-development and management expertise.

SciPhD's workshops introduce scientists to skills that correlate with success in entry-level industry positions: productivity, risk management and even empathy, for example. "Companies are not going to hire you because of technical expertise — rather, they want somebody innovative with good management skills who can play well in the sandbox," says Ribaud.

The workshops also help scientists to reframe their research experience on short, focused CVs, and in mock interviews. Juggling two or three sets of experiments and publishing papers, for example, could translate to "managed multiple projects simultaneously to meet milestones and time objectives".

Ribaud hopes that participants also leave with a sense of how many types of industry position they would qualify for with just a bit of added experience — jobs such as product manager, quality-control officer or analyst for a venture-capital firm.

Workshops are great for exposure to careers and for mini-injections of training. But nothing compares to the value of on-the-job experience, which students can gain from internships



Students learn techniques for prioritizing ideas at a Business of Science course run by SciPhD.

and volunteer opportunities. Watt, for example, participates in a student consulting group that partners with biotech start-ups to offer scientific advice at reduced consulting fees. Others gain experience by volunteering in their university's technology-transfer offices, which help to commercialize research discoveries.

One of the boldest and riskiest ways to gain real-world experience is to start a company based on a scientific discovery. This approach, experts say, is best after some formal business training and with hands-on mentoring from successful entrepreneurs (see 'Helping hands').

ON THE JOB

To ensure that its graduate students find a clear path to business careers and have opportunities to gain on-the-job experience, Johns Hopkins University School of Medicine in Baltimore, Maryland, established the Biomedical Careers Initiative (BCI). Peter Espenshade, a cell biologist, explains that the initiative grew out of a lack of career-development resources for trainees at a time when the job market was being squeezed.

Many graduate programmes have begun to showcase non-academic career options, but the BCI has taken steps beyond that. Its website curates and promotes non-science courses on campus, including those that help to build skills and knowledge in business.

The BCI also facilitates a three-month industry internship programme in which late-stage graduate students take a leave of absence from their research to work at companies or organizations. Students can do internships at companies with which the BCI has connections, or find their own opportunities. Two crucial components help to ensure buy-in from faculty advisers. One is that the BCI pays the student's graduate stipend during the internship; the other is a memorandum that all parties sign that documents the student's

interest, the start and end dates of the placement and an agreement between the student and adviser that certain research priorities must be completed first. This arrangement minimizes disruption to laboratory work.

In the BCI's first year, students completed internships at places such as MedImmune, Eli Lilly and the American Society for Biochemistry and Molecular Biology. Espenshade, who has trained a dozen PhD students in his own lab, says that he will count the BCI a success if it keeps even one student from "going reflexively into a postdoc as the next step". Ideally, graduate students would investigate non-academic career options from the beginning of their PhD programme, he says.

Gregory Cherryholmes did just that. Like Watt, he entered graduate school with an eye towards a biotech business career, and he completed the Keck certificate programme and did some freelance marketing consulting along the way. As a postdoc studying cancer immunotherapy at University of Washington in Seattle, he wants to help researchers in his field to ferry their cancer vaccines to the market.

He draws an analogy between most biomedical trainees' lack of business skills and a business concept called the 'valley of death' — the difficult stretch between having a good scientific idea and convincing investors to commercialize it.

"PhDs have all these skills — super bright with great critical thinking — but they don't have the business know-how to know where they want to go," he says. "These micro-opportunities can help bridge that valley." ■

Kendall Powell is a freelance writer in Lafayette, Colorado.

GRADUATE STUDIES

US education appeals

First-time enrolments of international students in US graduate schools rose for five consecutive years from 2009 to 2014, finds a report from the Council of Graduate Schools in Washington DC. The biggest year-on-year increases were in the physical and Earth sciences, up 20% to 16,235 first-time enrollees in 2014. The largest increase per nation was in the number of students from Brazil, which almost doubled its number of first-time US graduate-school enrolments to 1,134; India was up 27% to 21,889. However, first-time enrolments from Canada and China each fell by 1%, after previous annual increases of 3% and 5%, respectively. Jeff Allum, author of the report, says that the dip in enrolments from China might be a result of the nation's students taking advantage of their home country's investment in graduate schools. Meanwhile, a report from the Institute of International Education in New York found that overall international enrolments in US graduate schools rose by 6% to about 330,000 from 2013 to 2014. Nearly 80% of students from India were enrolled in science, technology, engineering and maths studies, as were 42% from China and one-fifth from the United Kingdom.

RESEARCH FUNDING

Wellcome change

The Wellcome Trust, the United Kingdom's largest biomedical-research charity, is rebalancing its funding priorities to focus on early-career scientists, collaborations and high-risk, high-reward projects. The London-based trust, which spends more than £700 million (US\$1.1 billion) a year on biomedical research and outreach, plans to boost funds for postdoctoral fellowships and to introduce small 'seed' grants to support research on innovative ideas for which few preliminary data are available. Separate programmes for new and senior investigators will merge, with candidates' career stages considered in grant-application evaluations — a move designed to favour early-career researchers. The shift comes after lengthy consultations with grant applicants, according to Jeremy Farrar, who took the reins at the foundation in April 2013. In a statement, he said that the new framework would help to channel more resources to the most promising questions. "We want to make sure that as we increase our funding, the right opportunities are available," he said. For an interview with Farrar, see *Nature* <http://doi.org/xjdk> (2014).

REVERSAL OF MISFORTUNE

Alternative culture.

BY J. W. ARMSTRONG

The philosophers claim that knowing you're going to be hanged in the morning focuses the mind wonderfully. This is not completely true. Also, I suspect that the philosophers do not speak from first-hand experience.

As I waited, an hour before my execution, my unfocused mind ranged freely — jumping from event to event — over how this had come to be.

It started five years ago. My colleagues and I were working to resolve a paradox in theoretical physics related to the many-worlds interpretation of quantum mechanics — the idea that the Universe bifurcates every time any decision is made, leading to a multiverse of alternate realities. Our breakthrough was, in retrospect, obvious. Without going into detail, we demonstrated that many-worlds is true, but with a twist: most decisions are unimportant and just blur the timeline. True bifurcation does not occur for *every* decision; rather the multiverse branches only on some decisions, perhaps, by definition, the 'consequential' ones.

But we went further. We showed that information leaks between nearby timelines. You can't visit a reality-that-might-have-been, but you can use its leaked information to glimpse (eavesdrop on?) the consequences of decisions you might have made differently. In addition to the physics, the personal possibilities were fascinating. Should you have married your first love? Should you have accepted that career-changing job offer? Should you have had children with the partner you did? Should you have had children at all?

We should have left it at that — establishing that leaked information from other timelines was detectable in principle. But I had an insight about how, in a practical way, the leaked information could actually be detected. We were almost giddy, and published quickly; I was the first author of Fossman *et al.*

Academic accolades followed immediately, including a very fast-track Nobel. The detector — the media dubbed it the 'Fossman device' — also went from theory to practice very quickly. At first the cost was enormous and the 'range' of the detector was severely limited by the

technology — only entities with the resources of a government could afford it, and only the nearest timelines could be detected. Militaries were interested, of course. Surprisingly to me, organized religions brought massive



resources to bear and were also early adopters. But the technology improved swiftly, the cost plummeted, and the number of alternate timelines accessible increased markedly. Soon corporations, presumably for some competitive advantage, used the technology. Then wealthy individuals could search their alternate timelines. Finally the technology was commodified, cell-phone-like, and was available to everyone.

When it got to that point — where any individual could afford to search thousands of alternate histories — it became an explosive social phenomenon.

The evidence mounted slowly at first. But as more and more people searched their alternate realities it became clear: the world we have fashioned — the Universe created by our collective decisions — was not 'average' compared with what could-have-been. No ... rather, our collective decisions were abysmally poor.

Not everyone's, of course — there were examples of people having made excellent life choices. But the average decisions individuals made in our Universe were

substantially inferior to those that could have been made — and those that actually were made in our alternate universes.

The initial public reaction was shock. TV chat shows, ostensibly trying to put our world in a multiverse perspective, trotted out philosophers, physicists and theologians to argue determinism, causality and free will in varying levels of sophistry. Late-night infomercials — for four easy payments of \$39.95 — flogged products and services to absolve guilt: your life decisions, bad though they might have been, were not your fault; rather you were a victim of randomness, malefic supernatural entities or whatever else would sell the product. Public shock morphed to sullen bewilderment — and then, quickly, to rage. People were angry and responded in the time-honoured human way: they looked for someone else to blame for their troubles. As the inventor of the technology that had revealed humanity's collective bad judgement, I was the obvious choice.

A mob mentality ensued and the government issued a Writ of Inquisition. I was tried and convicted of 'crimes against humanity'. The conviction was swift — perhaps because my defence team, too, had examined their alternate lives, hated themselves (and me for allowing them this knowledge) and put up only a token effort. I was sentenced to hang by the neck until dead.

Which brings us to now, minutes before my execution, and the focusing of minds ...

I'd read someplace that it is important to be stoic at such times. As I was marched to the gallows, seeing the hangman's noose, I nonetheless swallowed hard and had an uncontrollable desire to scratch my neck. I was saved from the ignominy of this gesture only because my hands were tied behind my back.

But in fact I can say, as I slowly mounted the scaffold, that I departed myself absolutely as well as could be hoped. You see I, too, had used the Fossman device, in this case to look at my endgame in other universes — and the alternatives there were very unpleasant indeed. As the noose was fitted around my neck, I took what comfort I could in knowing that I was about to depart the best of all possible worlds. ■

J. W. Armstrong works at a large laboratory in Southern California. The technology described here is fiction. Really. And if it isn't, none of this applies to us.

ILLUSTRATION BY JACEY

natureOUTLOOK

LIVER CANCER

4 December 2014 / Vol 516 / Issue No 7529



Cover art: Katie Scott

Editorial

Herb Brody,
Michelle Grayson,
Lauren Gravitz,
Kathryn Miller,
Lewis Packwood

Art & Design

Wesley Fernandes,
Denis Mallet,
Andrea Duffy

Production

Karl Smart,
Ian Pope,
Robert Sullivan

Sponsorship

Stella Yan,
Janet Cen,
Yvette Smith

Marketing

Hannah Phipps,
Elisabetta Benini

Project Manager

Anastasia Panoutsou

Art Director

Kelly Buckheit Krause

Publisher

Richard Hughes

Chief Magazine Editor

Rosie Mestel

Editor-in-Chief

Philip Campbell

Unless liver cancer is caught early, a patient's odds are dismayingly low. The disease hits hard and fast: survival rates vary by country, but in the United Kingdom, just 20% of people with the disease are still alive one year after diagnosis. It is also the sixth most common form of cancer worldwide, with nearly 800,000 cases reported in 2012. Yet despite liver cancer's growing prevalence, there is a startling lack of treatments. Beyond physical approaches such as radiation, transplant and surgery (page S17), there is only one approved therapy, which has been followed by a string of expensive failures and a slim list of candidate drugs (page S4).

Liver cancer is nearly always caused by underlying disease, but the source of that disease varies widely around the world (page S2) and includes hepatitis B virus (HBV), poor diet and inactivity, and fungal toxins. For each of these, public-health initiatives have the potential to make a difference: vaccines can prevent HBV (page S12), improved health practices can slow or even halt the progression of fatty liver disease (page S8), and better agricultural practices and storage conditions can limit fungal growth (page S7).

Researchers do not yet understand why some people can live with liver disease for many years, whereas others develop fatal cancer. They think that a sequence of biological hits must happen in order for liver disease to become malignant, but identifying these hits is proving challenging. Scientists are looking for clues far and wide, from the bacteria that live in our gastrointestinal tracts (page S14) to the gender differences that cause many more men than women to develop the disease (page S10).

We are pleased to acknowledge that this Outlook was produced with the support of Tongji Hospital. As always, *Nature* retains sole responsibility for all editorial content.

Lauren Gravitz
Contributing Editor

CONTENTS

S2 EPIDEMIOLOGY

A preventable cancer

The multiple causes of malignancy around the world

S4 DRUG DEVELOPMENT

Try and try again

High-profile drug failures highlight the struggle to find a cure

S7 PERSPECTIVE

Time to face the fungal threat

Felicia Wu charts the link between a toxin in food and liver cancer

S8 FATTY LIVER DISEASE

The liver labyrinth

Navigating the path between obesity and malignancy

S10 SEX DIFFERENCES

Luck of the chromosomes

More men than women are affected by liver cancer

S12 VACCINES

Taking a shot at protection

The hepatitis B vaccine is being taken up across the globe

S14 MICROBIOME

The bacterial tightrope

Imbalances in the microbiome and their role in liver cancer

S17 PERSPECTIVE

Incision revision

Myron Schwartz analyses how hepatitis drugs have affected liver transplants

COLLECTION

S18 Exploration of liver cancer genomes

T. Shibata & H. Aburatani

S28 *In vivo* RNAi screening identifies a mechanism of sorafenib resistance in liver cancer

R. Rudalska et al.

S37 Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma

A. G. Singal et al.

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the *Nature Outlook* Editorial guidelines available at go.nature.com/e4dwzvw

CITING THE OUTLOOK

Cite as a supplement to *Nature*, for example, *Nature* Vol. XXX, No. XXXX Suppl., Sxx–Sxx (2014).

VISIT THE OUTLOOK ONLINE

The *Nature Outlook Liver Cancer* supplement can be found at <http://www.nature.com/nature/outlook/liver-cancer>. It features all newly commissioned content as well as a selection of relevant previously published material.

All featured articles will be freely available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES

For UK/Europe (excluding Japan): Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas – including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea: Nature Publishing Group – Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

CUSTOMER SERVICES

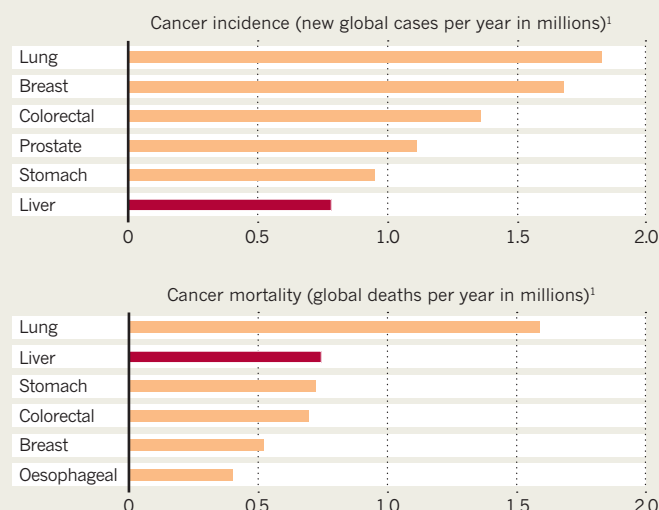
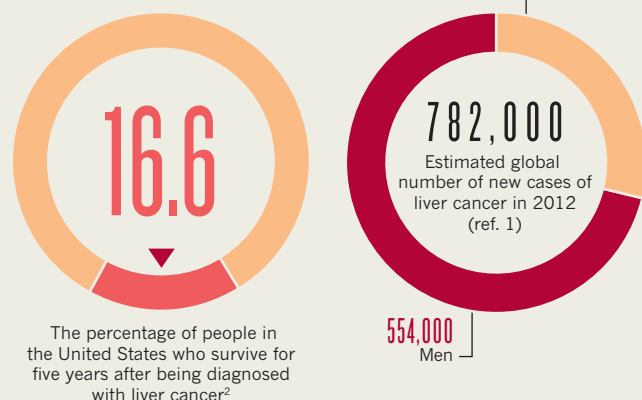
Feedback@nature.com
Copyright © 2014 Nature Publishing Group

A PREVENTABLE CANCER

Liver cancer is difficult to treat, and lethal if not caught early. But its most common causes, such as hepatitis viruses and obesity, can be prevented. By Lucas Laursen.

OUT OF PROPORTION

Not only does liver cancer strike more men than women, it also causes a disproportionate number of cancer-related deaths. Although it is the sixth most common malignancy worldwide, it is second only to lung cancer in terms of mortality¹.

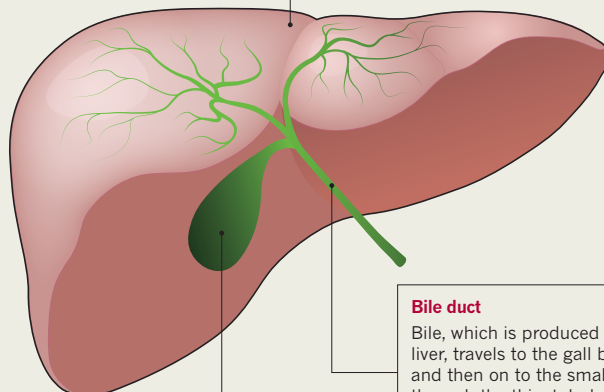


DOMINANT SUBTYPE

Liver cancer has several subtypes, but one — hepatocellular carcinoma (HCC) — is by far the most common worldwide. In some countries, the proportion of subtypes differs as a result of different risk factors.

Liver

The liver contributes to a wide range of functions, including digestion, detoxification and metabolism. It is also the only internal human organ that can regenerate: as little as 25% of its original tissue is necessary to restore the liver to its original size. HCC is named after the cells in which it develops, the hepatocytes.



Gall bladder

Bile acids, which are used in digestion, are stored in the gall bladder and released into the small intestine on ingestion of fatty foods. Certain bacteria in the gut convert bile acids into toxic chemicals that might contribute to liver cancer.

Bile duct

Bile, which is produced in the liver, travels to the gall bladder and then on to the small intestine through the thin, tubular bile duct. Bile-duct cancer, also known as cholangiocarcinoma, is less common than disease that starts in the lobes of the liver itself.

~78% HCC

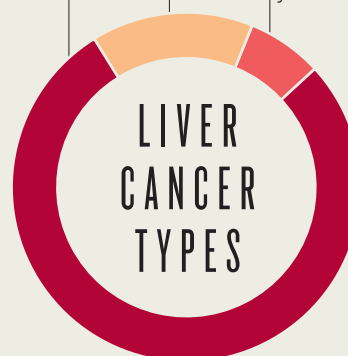
HCC accounts for the overwhelming majority of liver cancers³. It is the disease addressed by this Outlook, as well as by most liver-cancer research.

~15% Bile-duct cancer

Bile-duct cancer is more common than HCC in some Asian countries, but it makes up a relatively small number of liver-cancer cases worldwide⁴.

~7% Hepatoblastoma and various liver sarcomas and carcinomas

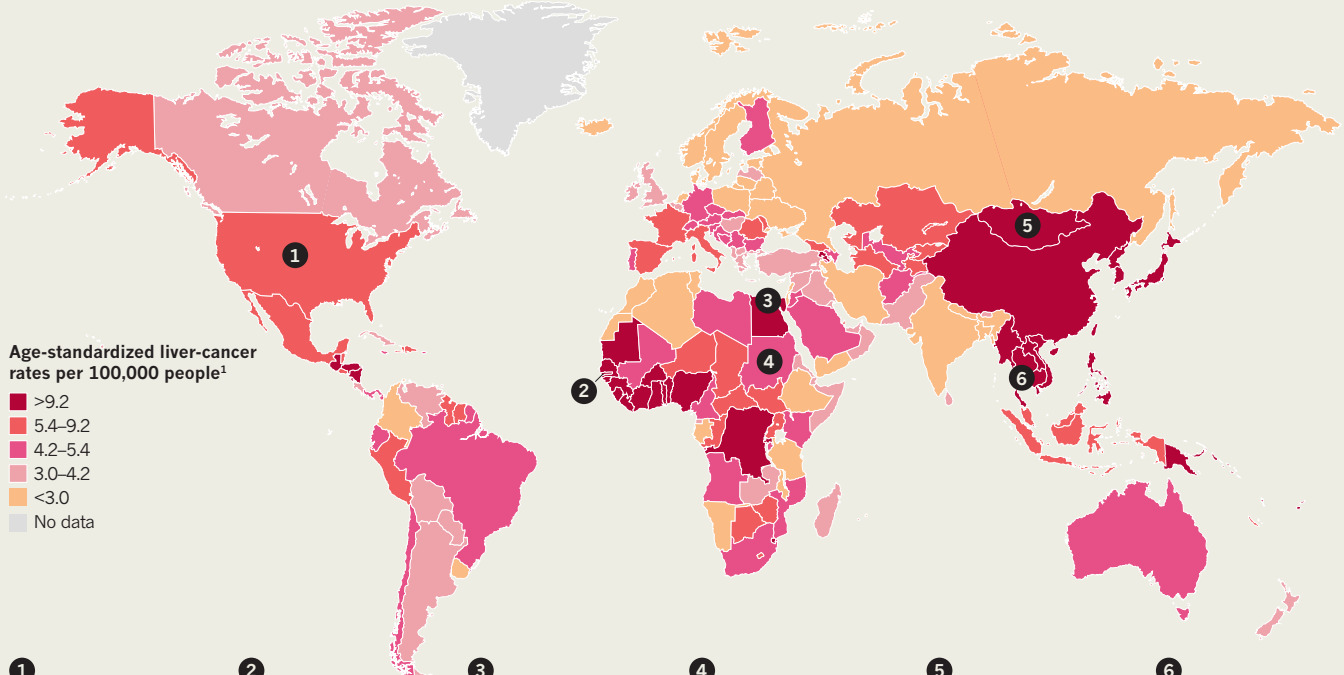
These include most cases of paediatric liver cancer, which has increased in incidence in recent years but is still a rare disease.



References: 1. *GloboCan 2012* <http://globocan.iarc.fr/>; 2. US National Cancer Institute Surveillance, Epidemiology, and End Results Program; 3. Ahmed, F. *et al. Prev. Chronic Dis.* **5**, A74 (2008); 4. Parkin, D. M., Ohshima, H., Srivatanakul, P. & Vatanasapt, V. *Cancer Epidemiol. Biomarkers Prev.* **2**, 537–544 (1993); 5. Larsson, S. C. & Wolk, A. *Br. J. Cancer* **97**, 1005–1008 (2007); 6. Chen, Y., Wang, X., Wang, J., Yan, Z. & Luo, J. *Eur. J. Cancer* **48**, 2137–2139 (2012); 7. Njei, B., Rotman, Y., Ditah, I. & Lim, J. K. *Hepatology* <http://dx.doi.org/10.1002/hep.27388> (2014); 8. Ananthakrishnan, A., Gogineni, V. & Saeian, K. *Semin. Intervent. Radiol.* **23**, 47–63 (2006);

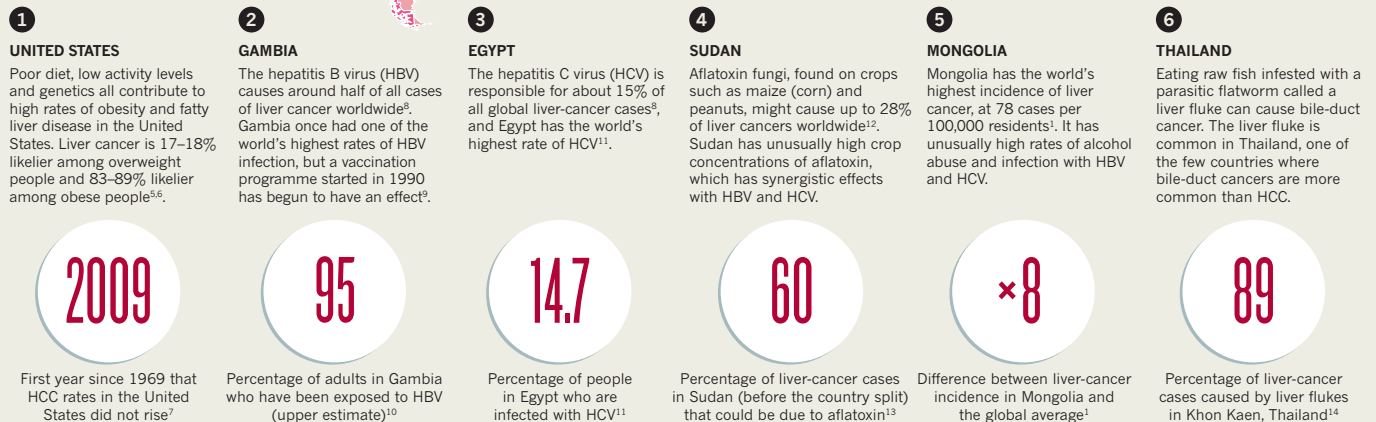
RISK ASSESSMENT

Liver-cancer rates vary widely between countries, with the highest rates occurring in Africa and east Asia. The disease is nearly always preceded by chronic liver damage, most commonly caused by hepatitis B. Highlighted below are other common causes in selected countries.



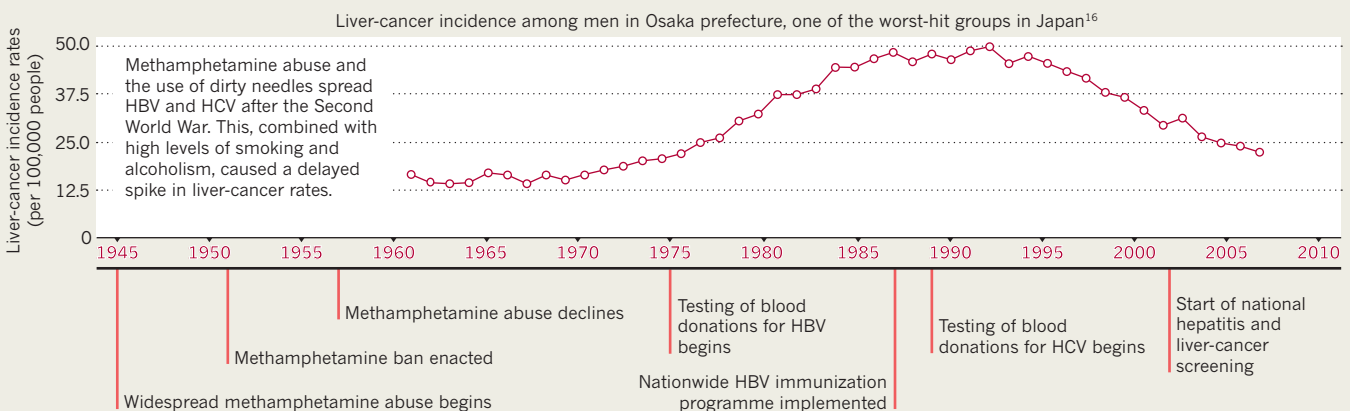
Age-standardized liver-cancer rates per 100,000 people¹

■ >9.2
■ 5.4–9.2
■ 4.2–5.4
■ 3.0–4.2
■ <3.0
■ No data

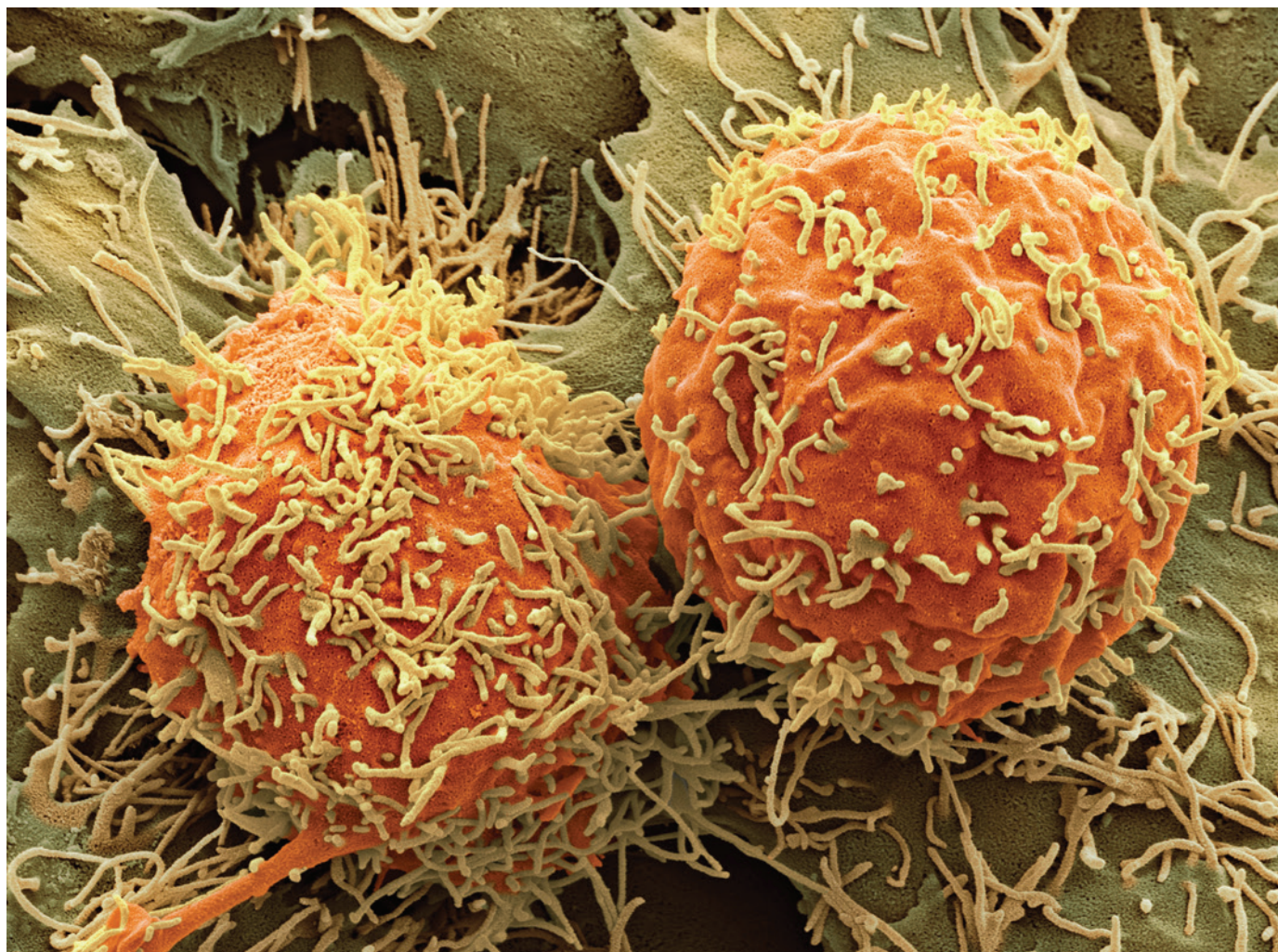


A SUCCESS STORY

Liver disease takes decades to progress to cancer, which makes prevention research difficult. In Japan, a 30-year gap separated the peak of hepatitis infections (both HBV and HCV) and the resulting rise in liver cancer¹⁵. But public-health efforts on multiple fronts are bringing liver-cancer rates back down.



9. Peto, T. J. et al. *BMC Infect. Dis.* **14**, 7 (2014); 10. Hwang, E. W. & Cheung, R. N. A. *J. Med. Sci.* **4**, 7–13 (2011); 11. Mohamoud, Y. A., Mumtaz, G. R., Riome, S., Miller, D. & Abu-Raddad, L. J. *BMC Infect. Dis.* **13**, 288 (2013); 12. Liu, Y. & Wu, F. *Environ. Health Perspect.* **118**, 818–824 (2010); 13. Omer, R. E. et al. *Nutr. Cancer* **48**, 15–21 (2004); 14. Vatanasapt, V., Sriamporn, S. & Vatanasapt, P. *Jpn J. Clin. Oncol.* **32**, S82–S91 (2002); 15. Gersten, O. & Wilmoth, J. R. *Demographic Res.* **7**, 271–306 (2002); 16. *CI5plus* <http://ci5.iarc.fr/CI5plus>.



STEVE GSCHMEISSNER/SPL

A coloured scanning electron micrograph of two hepatocellular carcinoma cells. Patients with the disease typically live for only 6 to 20 months after diagnosis.

DRUG DEVELOPMENT

Try and try again

Drug companies have been fighting a losing battle against advanced liver cancer — but refinements of proven techniques along with radical new approaches could turn the tide.

BY MEGAN SCUDELLARI

At a subdued Friday afternoon poster session at a small conference in San Francisco, California, in January 2013, scientists from the pharmaceutical company AbbVie, based in Chicago, Illinois, quietly presented the findings from a phase III clinical trial for a highly anticipated liver-cancer drug. The drug had failed. Tested in more than 1,000 patients over 3 years, it offered no significant difference in survival over the standard drug therapy. The disappointment was palpable.

AbbVie had produced the latest lead balloon in a string of expensive, spectacular failures of advanced liver-cancer drugs: a year earlier, a

therapy by the New York City-based pharma company Bristol-Myers Squibb had floundered in phase III, preceded by a 2011 flop from the drug giant Pfizer (also based in New York City). And this summer, the blows kept coming. In June, Eli Lilly, a pharma company headquartered in Indianapolis, Indiana, announced a phase III miss. Most recently, the Swiss drug company Novartis Pharmaceuticals, based in Basel, chalked up a failure in July. All told, five large, expensive phase III trials have failed in the past four years. “It’s a bit of a graveyard right now,” says Paul Lammers, the president and chief executive of Mirna Therapeutics, a biotech company based in Austin, Texas.

All five drugs were designed to treat

hepatocellular carcinoma (HCC), which makes up 70–85% of all cases of liver cancer in most countries. HCC can be cured with surgery or a liver transplant if caught promptly, but only 20–30% of HCC cases are detected in the early stages. For most patients, the cancer is too advanced or underlying conditions, such as hepatitis or fatty liver disease, make the liver too fragile for invasive treatments. And without them, HCC is quick and deadly: patients typically live for only 6 to 20 months after diagnosis. According to data from the National Cancer Institute, in the United States just 15% of those with advanced disease survive one year after being diagnosed, highlighting the dire need for new therapies. So many recent

clinical-trial failures have brought a growing sense of urgency, and researchers are pushing hard to investigate therapeutic strategies that break new ground as well as those that borrow heavily from existing approaches.

Today, only one drug is approved to treat advanced HCC. Sorafenib, an oral medication marketed as Nexavar by the drug companies Bayer (Leverkusen, Germany) and Onyx Pharmaceuticals (San Francisco, California), is thought to work by binding to and shutting down key receptor tyrosine-kinase enzymes in two biological pathways that are implicated in cancer. When active, these kinases promote cell replication and the growth of new blood vessels, which are crucial for tumour survival and expansion. Sorafenib can buy patients time, but the drug is far from ideal: treatment costs US\$5,400 per month, but it extends lifespan by an average of only 2.8 months and causes a range of side effects that include diarrhoea, nausea, fatigue and skin reactions.

When sorafenib was approved for HCC in 2007, many heralded it as the beginning of a new era of targeted HCC therapies, and look-alikes quickly appeared in pharmaceutical pipelines. Three of the four drugs in sorafenib's wake — AbbVie's linifanib, Bristol-Myers Squibb's brivanib and Pfizer's sunitinib — also inhibit multiple tyrosine kinases. Novartis's everolimus works in a similar fashion by blocking a protein called mTOR that activates the cell cycle and blood vessel growth. None worked better than sorafenib. Researchers also tried supplementing sorafenib with additional kinase inhibitors, such as erlotinib (developed by Genentech in San Francisco, California), but there was no improvement.

After the approval of sorafenib, companies expected to improve on it quickly and produce many more drugs for HCC that capitalized on the same mechanism. "People thought we'd be done tomorrow," says Jordi Bruix, head of the Oncology Liver Unit of the Hospital Clínic de Barcelona at the University of Barcelona in Spain, who designed and ran the phase III clinical trial of sorafenib. "This has all failed."

"There was a lot of expectation put into molecular medicine, but this has not been fulfilled."

Today, some drug developers wonder whether sorafenib even works in the way they originally thought. "Maybe we got tricked. While sorafenib hit that tyrosine-kinase pathway, it may not be the key mechanism by which it's working," says David Kirn, executive chairman of the biotech company SillaJen (based in Busan, South Korea), which is pursuing an HCC drug therapy. This leads to the next question, he says: what else is sorafenib hitting? It is a question that does not yet have an answer.

Some companies are continuing to target HCC with kinase inhibitors, but only in a



David Brown works on drug candidate MRX34 at the Mirna Therapeutics laboratory in Austin, Texas.

subset of people with specific biomarkers that are thought to identify patients who will benefit most from that approach¹. Other firms are beginning to look beyond sorafenib to new targets and new ways to hit them, such as by using RNA interference (RNAi) technology or cancer-killing viruses, or by attacking the cancer before it digs in.

DIFFICULT DEVELOPMENTS

Molecular medicine — the identification and targeting of specific genetic errors and molecular changes that cause disease — has been highly successful in improving the understanding and treatment of many types of cancer. But liver cancer is not one of them². The causes and progression of HCC are obscured by labyrinthine complexity.

There are three primary complexities. First, the disease is extremely diverse in both its causes and genetics. HCC develops from cirrhosis (that is, scarring of the liver due to chronic liver disease) in more than 90% of cases; but cirrhosis can result from any number of conditions, such as alcoholism, infection with hepatitis B or C, or the build-up of fat in the liver. "These patients have two major diseases; they have cancer, and they have major liver dysfunction," says Kirn.

Furthermore, whereas most cancers are driven by certain key genetic mutations, the mutations associated with HCC vary widely among patients and even within a single tumour — two different regions in the same tumour can have strikingly different mutations. Because of this dramatic diversity, finding a reliable drug is like trying to hit a moving target.

This variation also precludes the use of genetic testing to categorize patients for treatment: there are no clear genetic signatures by which to stratify them. "All the [genetic] classifications are useless to treat patients," says Bruix, whose research team developed the most widely used system for staging and treating HCC.

"There was a lot of expectation put in molecular medicine, but this has not been fulfilled."

The second challenge for drug developers is that liver cells are a fortress against drugs. Liver-cancer cells show enhanced expression of proteins that confer drug resistance, and liver cancer is regularly resistant to chemotherapy. Moreover, attempting to overcome that resistance with combinations of drugs is risky: treatments that have any degree of toxicity can exacerbate the underlying liver disease.

Finally, patients with HCC who are eligible to enrol in phase III clinical trials — the true testing ground for a drug's efficacy — are typically very ill. To a large extent, this is the nature of cancer-drug development: phase III clinical trials generally enrol patients who have failed to respond to approved therapies and are in the advanced stages of disease. Therefore, "the chance to see a dramatic impact is very slim," Lammers says.

This is particularly evident in liver cancer, in which late-stage patients deteriorate quickly as their livers begin to fail. In fact, companies struggle to enrol sufficient numbers of patients into HCC clinical trials because many individuals do not have a long enough life expectancy to qualify for the trial.

A MIX OF OLD AND NEW

Today, the pharma industry is split between two camps with very different drug-development philosophies for advanced liver cancer: those who are continuing to pursue kinase inhibitors such as sorafenib, and those who are attempting to hit new drug targets using novel strategies (see 'Clinical trials for advanced hepatocellular carcinoma').

Industry veteran and physician Brian Schwartz, who led the initial development of sorafenib at Bayer, thinks that additional kinase inhibitors can succeed. Companies just have to be more careful about which patients receive which drugs, he says.

Now at the biotech company ArQule in Woburn, Massachusetts, Schwartz is overseeing a phase III trial of tivantinib. Tivantinib is a small-molecule inhibitor of c-MET, which is a receptor tyrosine kinase involved in cell proliferation and blood vessel growth. With the aim of boosting the drug's chance of success, the company is enrolling only patients who have high levels of c-MET on the surface of their tumour cells. Schwartz expects that such patient stratification will become the norm for HCC, as it is in other cancer fields. The recent failed trials might have got closer to success had they stratified them differently, he says. "That's why we spent a lot of time making sure we would stratify this trial correctly."

But other companies think that a more audacious approach is needed. Tekmira Pharmaceuticals, based in Burnaby, Canada, develops RNAi-based drug candidates. RNAi silences specific disease proteins by chopping up those proteins' messenger RNAs: molecules that carry

CLINICAL TRIALS FOR ADVANCED HEPATOCELLULAR CARCINOMA

Several prominent liver-cancer drug candidates have failed in recent years, but here are some therapies that are in clinical trials.

Treatment (company)	Trial phase	ClinicalTrials.gov identifier	Notes
Receptor tyrosine-kinase inhibitors			
Axitinib (Pfizer)	II	NCT01210495	Already approved for advanced kidney cancer
Lenvatinib (Eisai)	III	NCT01761266	Inhibits multiple receptor tyrosine kinases
Regorafenib (Bayer)	III	NCT01774344	Currently approved to treat colorectal cancer; inhibits multiple kinases
Tivantinib (ArQule)	III	NCT02029157	Inhibits a cell receptor called c-MET; might be most effective against patients with high levels of c-MET on their tumours
Other targets			
Cixutumumab (US National Cancer Institute)	II (completed)	NCT00639509	A monoclonal antibody that blocks a protein for cell growth
LY2157299 (Eli Lilly)	II	NCT02178358	Inhibits signalling by transforming growth factor- β
MRX34 (Mirna Therapeutics)	I	NCT01829971	The first microRNA to enter clinical trials for hepatocellular carcinoma
Pexa-Vec (JX-594) (SillaJen)	II (completed)	NCT00554372	Activates the immune system using an engineered vaccinia virus
TKM-PLK1 (Tekmira)	I/II	NCT02191878	Silences a protein involved in tumour-cell proliferation

genetic information for a specific protein from the nucleus to the cytoplasm. With RNAi, a small piece of RNA is delivered to a tumour cell, where it works with a protein complex to disable specific pieces of mRNA. In this way, a chosen mRNA — the product of a cancer-causing gene, for instance — can be targeted and destroyed without affecting other processes in the cell.

Mark Murray, president and chief executive of Tekmira, wanted to apply RNAi technology to cancer, so the company screened an array of oncology targets to see whether any could be effectively silenced with RNAi.

One target rose like cream to the top: polo-like kinase 1 (PLK1), a protein that is involved in tumour-cell proliferation and is overproduced in liver-cancer cells. Other companies have tried to target PLK1 with small-molecule inhibitors, but these drugs travelled to the bone marrow and interrupted normal cell division, causing dangerous side effects. To avoid this, Tekmira enclosed its interfering RNA in a small bubble of lipids that accumulate only in the liver. “We’ve learned from previous attempts to inhibit PLK1 and built a better mousetrap, if you will,” says Murray. The final drug candidate, TKM-PLK1, is currently in a phase I/II safety and efficacy trial that is expected to conclude next year.

A different type of small RNA, microRNA, is also under investigation. In 2002, biochemist David Brown (then at Ambion in Foster City, California) and his colleagues began experimenting with these tiny strands of RNA, which do not carry instructions for creating specific proteins as mRNAs do, but instead directly coordinate activities in cells. Brown compared healthy human cells with tumour cells and identified a set of 20 microRNAs that are involved in cancer: 6 that are overabundant in cancer cells and 14 that are found at reduced levels or are absent in cancer cells.

One of the missing microRNAs, a 23-nucleotide snippet called miR-34, blocked the activity of some 25 oncogenes. Brown, now at Mirna

Therapeutics, and his team have now designed a synthetic version of miR-34. When they injected it into animal models of liver cancer, the drug shrank tumours until they vanished. Today, Mirna’s drug is in a phase I safety trial and is the first microRNA to enter clinical trials for HCC.

Others are investigating whether the growing field of cancer immunotherapy, which uses the body’s own immune system to attack the disease, might be a better fit for tackling liver cancer. In March, SillaJen acquired the San Francisco-based biotherapeutics firm Jennerex, which has been attempting to use a virus to direct the immune system against liver-tumour cells. The company’s drug, Pexa-Vec, is an engineered vaccinia virus — the same virus that is the main component in the smallpox vaccine. The virus preferentially infects cancer cells because it is activated only by certain enzymes produced by tumours. On activation, the virus multiplies inside tumour cells and causes them to burst open. In the process, it releases antigens that attract immune-system cells to the tumour site. The activated virus also attracts adaptive immune cells, using both its own viral proteins and proteins generated by a synthetic gene inserted into the virus.

In a randomized phase II trial of 30 patients with advanced liver cancer, high doses of Pexa-Vec resulted in a median survival time of more than 14 months: twice the survival time for patients on placebo³. A phase III trial is slated to begin next year in partnership with the biotech company Transgene (based in Strasbourg, France) and Lee’s Pharmaceutical (based in Hong Kong).

REMOVING THE ROOTS

Another small yet vocal contingent of academic researchers advocates targeting HCC much earlier than today’s drug candidates do, before the disease progresses.

One tantalizing target is the protein that seems to cause HCC in people infected with hepatitis. At Temple University in

Philadelphia, Pennsylvania, Mark Feitelson has spent the past 30 years studying how hepatitis B leads to liver cancer. He found that the hepatitis B virus causes cancer through the action of a small protein known as hepatitis Bx (HBx). HBx is one of four proteins produced by the hepatitis B genome and is the only one involved in the regulation of viral gene expression (the other three are structural). It also seems to be a cancer-causing brute: HBx binds to and inactivates tumour suppressors, upregulates oncogenes and blocks the immune system from killing viral-infected cells.

If a drug could target HBx in hepatitis B-infected livers, Feitelson says, it could pull out the cancer by its roots, instead of waiting until the cancer was established and then trying to simply trim its branches. “If someone gets a cut, do you wait until it is infected and amputate the arm?” he asks. “Or do you put on a band aid with a topical antibiotic and be done with it?” Targeting HCC early could also help to limit the genetic variety that develops over time, resulting in more-uniform cancer cells that are easier to identify and kill.

Enough early markers for liver cancer exist — such as hepatitis infection and cirrhosis — that companies should attempt to use drugs far earlier than they do, Feitelson argues.

For now, however, drug development in HCC remains focused on late-stage cancer and the incremental steps that could extend patients’ lives beyond the extra three months offered by sorafenib. “If we can make [HCC] into a more manageable disease with a quality of life that fits, that would be great,” Lammers says. “We have high hopes, but nobody is talking about a cure.” ■

Megan Scudellari is a freelance science writer in Boston, Massachusetts.

1. Finn, R. S. *J. Hepatol.* **60**, 243–244 (2014).
2. Bruix, J., Gores, G. J. & Mazzaferro, V. *Gut* **63**, 844–855 (2014).
3. Heo, J. *et al. Nature Med.* **19**, 329–336 (2013).

PERSPECTIVE



Time to face the fungal threat

Changing crop selection and improving food storage might reduce global rates of liver cancer, says **Felicia Wu**.

Moulds that grow naturally on food can produce toxins that have serious effects on health, even causing cancer. Certain moulds — especially those that grow in maize (corn), peanuts and tree nuts such as almonds and pistachios — produce aflatoxin, a carcinogen that is estimated to cause up to 28% of the total worldwide cases of hepatocellular carcinoma (HCC), the most common form of liver cancer^{1,2}. Many people are unaware of the link between naturally occurring food contaminants and cancer. Education, stricter food-quality controls and the cultivation of alternative crops can have a large impact on worldwide rates of liver cancer.

Aflatoxin is generated by the fungi *Aspergillus flavus* and *Aspergillus parasiticus* and, in addition to liver cancer, it has been associated with acute poisoning, immune-system dysfunction and stunted growth in children. People whose livers are already compromised by infection with hepatitis B virus (HBV) are particularly susceptible to aflatoxin-induced liver cancer³.

Studies by my team^{1,2} suggest that up to 172,000 cases of HCC per year can be attributed to exposure to aflatoxin in the diet, and most of these individuals are infected with HBV. The majority of cases occur in sub-Saharan Africa, southeast Asia and the western Pacific region (including China), as well as in parts of Central America.

TOXIN CONTROL

Although around 100 nations have set standards for the maximum amount of aflatoxin allowable in foodstuffs, setting a standard does not always ensure a safe food supply — particularly among subsistence farmers and rural communities in low-income countries, where food rarely undergoes formal safety inspection. Unfortunately, these countries also often have the highest levels of aflatoxin in their crops. And adding to the complexity is the fact that different nations consume different amounts of the riskiest crops. The United States and Kenya, for example, both have maximum allowable aflatoxin levels of 20 micrograms per kilogram of maize, but the average Kenyan consumes six times more maize per day than the average American, and usually in a less-processed form.

Aflatoxin-induced liver cancer can be prevented, but doing so will take a coordinated effort. One part of the solution is widespread vaccination against HBV, which would rob the toxin of its partner in carcinogenicity⁴ (see page S12). Beyond immunization, decreasing contamination can be implemented with little cost to farmers. Stressed plants are more vulnerable to fungal infection, so the best prevention is a healthy field of crops. This can be fostered in numerous ways, including the careful selection of crop varieties that are matched to the soil and water conditions, improved irrigation practices to reduce drought stress, and by breeding crops for both high yield and the ability to withstand environmental stressors, such as heat, drought and insect damage. Beyond stress prevention, farmers can use fungi that do not produce aflatoxin to compete with the *Aspergillus* species and prevent them from colonizing crops⁵.

Growing the food is just part of the problem, however: aflatoxin also accumulates in crops after they have been harvested. The fungi that produce aflatoxins thrive in damp conditions, so storage facilities must be kept cool and dry. Pests such as rodents and insects must be controlled because they can transport the fungus to other stored foods⁶. If these tactics fail and the food gets contaminated, it can still be salvaged: clays or chlorophyllin (chlorophyll-derived compounds) can be added to the food, either before or after distribution to consumers, to temporarily sequester aflatoxin in the gastrointestinal tract^{7,8}.

The simplest solution is to reduce consumption of foods that are contaminated with aflatoxin. Dietary changes in Qidong, China, show just how effective this can be. Market reforms introduced in the 1980s meant that citizens in the area began eating more rice and less maize. Decreased maize consumption (from around 100 kilograms per person in the 1970s to almost none in 2012) resulted in lower exposure to

aflatoxin in the diet and a 45% decrease in liver-cancer mortality⁹. This effect is independent of China's HBV vaccination programme, which began in 2002.

Qidong provides evidence that introducing dietary diversity can reduce the effects of aflatoxin-related liver cancer. But diverse diets serve another purpose: the various essential amino acids and other compounds found in leafy green, cruciferous and *Allium* vegetables can help to nullify dietary toxins¹⁰.

Now, it is crucial to disseminate these interventions and encourage those populations at the highest risk to adopt them. Locally, educators must encourage communities to grow and consume foods that are less likely to be infected with *Aspergillus*. But because dietary change occurs slowly, they should also help communities to

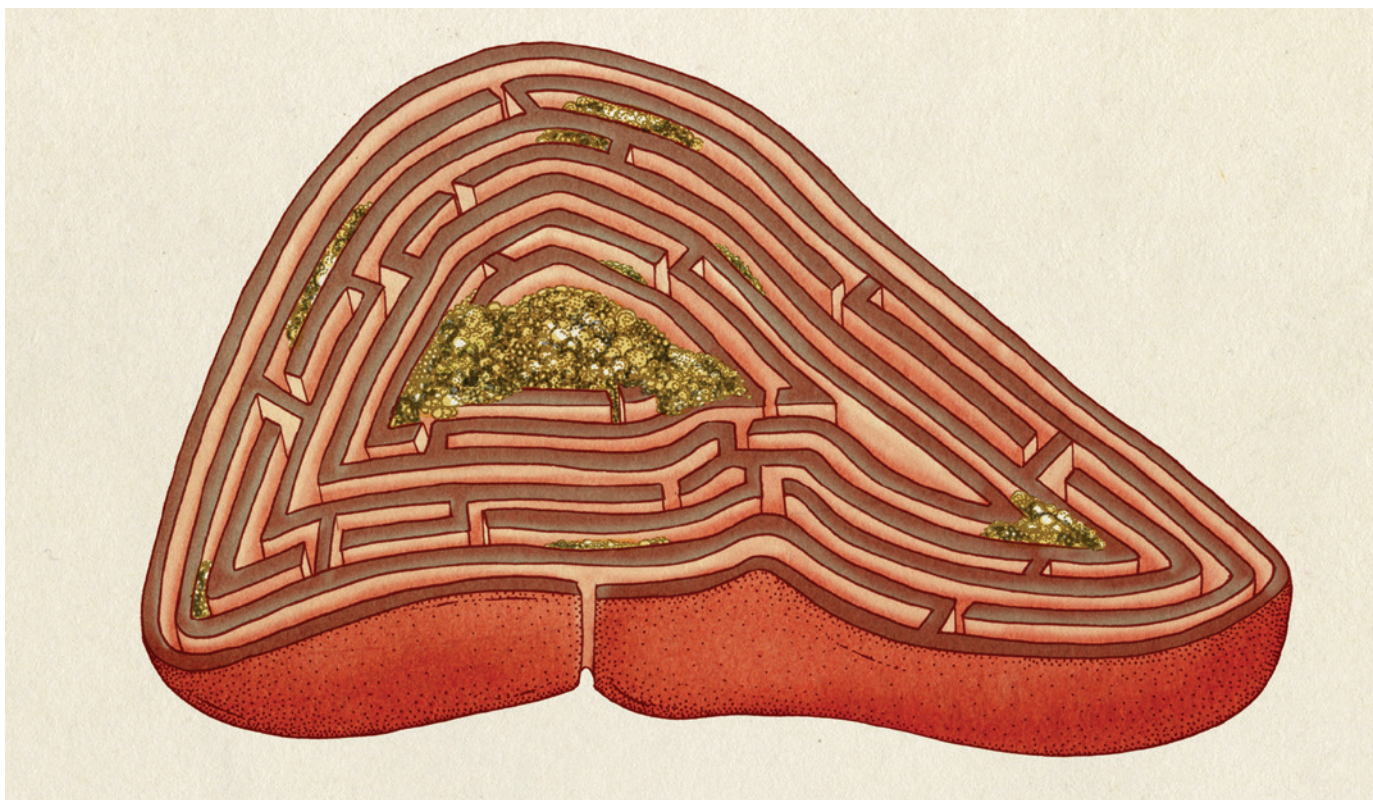
choose geographically suitable maize and peanut varieties and teach them practices that minimize post-harvest aflatoxin accumulation.

Globally, the HBV vaccine must be disseminated to high-risk populations at increased risk of aflatoxin exposure. And as the prevalence of hepatitis C virus (HCV) increases worldwide, researchers must investigate the potential synergism of HCV and aflatoxin exposure. Only then will rates of these preventable cases of liver cancer begin to decline. ■

Felicia Wu is a food, nutrition and agricultural economics professor at Michigan State University, East Lansing.

1. Liu, Y. & Wu, F. *Environ. Health Perspect.* **118**, 818–824 (2010).
2. Liu, Y., Chang, C. C., Marsh, G. M. & Wu, F. *Eur. J. Cancer* **48**, 2125–2136 (2012).
3. Groopman, J. D., Kensler, T. W. & Wild, C. P. *Annu. Rev. Public Health* **29**, 187–203 (2008).
4. Khlangwiset, P. & Wu, F. *Food Addit. Contam.* **27**, 998–1014 (2010).
5. Cleveland, T. E., Dowd, P. F., Desjardins, A. E., Bhatnagar, D. & Cotty, P. J. *Pest Manag. Sci.* **59**, 629–642 (2003).
6. Turner, P. C. *et al. Lancet* **365**, 1950–1956 (2005).
7. Mitchell, N. J. *et al. Am. J. Trop. Med. Hyg.* **91**, 777–785 (2014).
8. Egner, P. A., Muñoz, A. & Kensler, T. W. *Mutat. Res.* **523–524**, 209–216 (2003).
9. Chen, J.-G. *et al. Cancer Prev. Res.* **6**, 1038–1045 (2013).
10. Wu, F., Mitchell, N. J., Male, D. & Kensler, T. W. *Toxicol. Sci.* **141**, 329–334 (2014).

**AFLATOXIN-INDUCED
LIVER CANCER
CAN BE
PREVENTED,
BUT DOING SO
WILL TAKE A
COORDINATED
EFFORT.**



FATTY LIVER DISEASE

The liver labyrinth

Navigating the path that connects fatty liver disease to cancer has proved tricky.

BY BRANWEN MORGAN

The slimy-smooth, dark-red exterior of a healthy liver yields few clues about its complex nature. This 1.5-kilogram lump of tissue is so good at recovering from injury that it requires as little as one-quarter of its original bulk to fully regenerate. But the organ has its limits. Constant cell damage, such as that caused by a fatty diet, necessitates constant repair, providing the perfect setting for malignancy. Researchers are beginning to unravel how the transition to cancer occurs, with the aim of, eventually, finding more-effective ways to prevent it.

The body needs fat for hormone production, vitamin transport and insulation, and the liver stores and breaks it down in a tightly regulated process. The accumulation of too much fat, however, can result in 'fatty liver disease', and in some people this can develop into a condition called non-alcoholic steatohepatitis (NASH), in which sections of the liver become inflamed. NASH can be accompanied by scar tissue (fibrosis) that becomes thick and hard (cirrhosis), and a few cases (2–5% each year) progress to the most common form of liver

cancer, hepatocellular carcinoma (HCC).

The primary causes of cirrhosis are still alcoholism and hepatitis infection (see page S12), but NASH-based cirrhosis cases are rising fast. "The prevalence of fatty liver disease has doubled during the last 20 years, which is perhaps not surprising given the diabetes and obesity epidemic," says Geoff McCaughan, director of the Australian National Liver Transplant Unit at the Royal Prince Alfred Hospital in Sydney.

Even more troubling than the rise in NASH is another trend that McCaughan and others have identified. "Fifteen per cent of our patients with fatty-liver-associated cancers don't have cirrhosis," he says. The causes of these cases are unclear. Deciphering the patterns underlying these conditions has therefore become one of the most pressing challenges for those studying the origin and prevention of liver cancer.

As many as one in three people in developed nations might have fatty liver disease, according to the World Gastroenterology Organisation, and the condition is one of the top causes of liver disease in Western countries. But it is

unknown what pushes some of those people down the path to NASH, and others even further towards HCC. "There's a very strong link between NASH and HCC, but we don't yet have large studies with the level of confidence that we have for other aetiologies," says Lars Zender, a cancer researcher at the University of Tübingen, Germany. "We need data from large, prospective, epidemiological studies that follow patients with fatty liver disease through to tumour development."

Such studies, however, can take decades and require huge amounts of money and resources, and just diagnosing fatty liver disease in the first place is difficult. Abnormal blood tests can indicate that someone might have it, but the only accurate method for diagnosis is a liver biopsy — a complicated, invasive procedure that is still subject to sampling error.

Researchers assess the amount of fat in the liver on a scale from 0 to 3: grade 0 denotes a healthy liver in which less than 5% of liver cells, or hepatocytes, contain fat; grade 1 indicates mild fatty liver disease (5–33% of hepatocytes contain fat); grade 2 is moderate disease (33–66%); and grade 3 is severe disease (more than 66%). The higher the score, the higher the

GO NATURE.COM
For more on non-alcoholic fatty liver disease, see:
go.nature.com/wp3q53

likelihood of inflammation and cancer.

But the development of NASH depends on more than just the accumulation of a specific percentage of fat. “There is a tipping point, but it is different for different people,” says Helen Reeves, a hepatologist and cancer researcher at Newcastle University, UK. “There are people with a score of 3 who have no inflammation and people with a score of 1 who can have quite aggressive inflammation.”

THE INFLAMMATORY TRIGGER

Researchers have long known that it takes more than just fatty hepatocytes to trigger an inflammatory response and a subsequent transition to cancer. But they are only just beginning to understand what happens at a molecular level. New studies suggest that signalling molecules in and around hepatocytes determine whether prolonged high fat levels will become harmful. These molecules are produced by various tissues, as well as by bacteria that live in the gut, and they can be transported to the liver through the blood (see page S14). Furthermore, fat cells under the skin and around internal organs also release some of the signalling molecules that promote inflammation and subsequent liver damage.

These signalling molecules do not just target hepatocytes — they can also affect a different set of liver cells called hepatic stellate cells. Stellate cells produce proteins that act as a scaffold for the liver; when overactive, they deposit excessive amounts of collagen fibre in the spaces between hepatocytes, thus restricting blood flow and placing further stress on the liver.

Hepatocytes extract fat that is circulating in the bloodstream, then repackage and re-release it into the blood, store it as droplets inside the cell or break it down and use it for energy. High hormone and fat levels throw some of these processes into overdrive, and the faster they occur, the greater the accumulation of by-products, especially the reactive oxygen species commonly known as free radicals.

Free radicals react with and bind to many cell structures, preventing normal activity and putting a cell under oxidative stress. This stress causes hepatocytes to switch on self-destruction processes and activates local immune cells, exacerbating the inflammation. “If you have inflammation and oxidative stress, you are damaging your liver all the time,” Reeves says.

This sustained assault turns the liver’s capacity for repair, once its saving grace, into its undoing. Continuous fat uptake means that the liver is constantly engaged in a cycle of damage, inflammation and cell death. And the

constant cell division as the organ attempts to regenerate drastically increases the likelihood of cancer-causing mistakes being made during DNA replication.

DAMAGE CONTROL

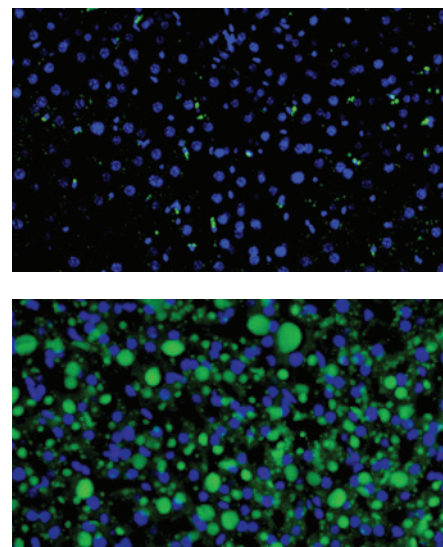
A better understanding of how cancer emerges from NASH requires a better monitoring system — one that does not rely solely on detecting cirrhosis. “The problem is, if NASH hasn’t advanced to cirrhosis, nobody is looking for cancer,” Reeves says.

Beyond that, researchers also need to pay more attention to the molecular profiles of diseased livers, says Zender. Examining tissue samples from affected and unaffected sections of someone’s liver could show which genes or genetic pathways are triggering or perpetuating abnormal cell growth. Comparing the data with that of other patients with NASH could also reveal what cell-death-promoting molecular pathways are being activated in the presence and absence of cirrhosis. Such insight could help researchers to find future therapeutic targets against NASH, with the hope that treating the condition early will prevent progression to cancer.

Finding a way to treat NASH should have benefits beyond cancer prevention. Indeed, there is a strong association between NASH and cardiovascular disease, so a treatment for NASH could reduce the mortality due to this associated condition. Eventually, many patients with NASH require surgical removal of part of their liver or a liver transplant (see page S17), so treating the disease early could reduce the number of operations and thereby lessen the need for scarce donor organs.

The past few years have seen a substantial change in the NASH therapeutic landscape. The lack of animal models has stymied the development of new treatments for fat-related liver cancer¹ because normal mice typically develop liver cancer without fibrosis or cirrhosis as intermediary steps. But new mouse models could remove this hurdle. Cancer researcher Michael Karin at the University of California, San Diego, recently described a mouse model of NASH that progresses to human-like HCC in 40 weeks². He and his colleagues then showed that interfering with a key inflammatory protein prevented NASH and the progression to liver cancer. The protein, called tumour-necrosis factor, is already targeted by a psoriasis and rheumatoid arthritis drug called etanercept; when Karin gave the drug to his mice, it prevented NASH and HCC. He hopes to convince companies that make anti-inflammatory drugs such as etanercept to test them in patients with NASH and NASH-driven HCC — failing that, he will organize such tests himself.

Thomas Burris, a pharmacologist at St Louis University, Missouri, feeds normal mice a diet similar to that of people in Western countries: full of saturated fat and cholesterol and high in



Lipids (green) in liver tissue from normal mice (top) and mice with fatty liver disease (bottom).

sugar in the form of fructose (which is abundant in soft drinks). He then studies how the animals progress from mild to severe fatty liver disease to NASH. The mice develop a fatty liver within weeks, with NASH emerging over a couple of months. A number of the older NASH mice even progress to HCC.

Burris’s mouse models have been used to find several therapeutic targets for the treatment of NASH, as well as drug candidates. One of these candidates, SR9328, has been shown to suppress fat production, eliminate inflammation and reverse fat accumulation in mouse livers³. But researchers do not yet know how effective this drug will be in the presence of fibrosis and cirrhosis, which do not develop in these mice. A phase II clinical trial of a drug with a similar mode of action — obeticholic acid — showed the medication to be so effective that its developer, Intercept Pharmaceuticals, based in San Diego, California, halted the trial early.

The need for therapies such as this is dire: current global projections suggest that in some countries NASH will overtake hepatitis as the main need for liver transplantation by 2025. The increase in NASH is also likely to lead to a rise in rates of HCC, but researchers are still unsure of the mechanisms — genetic or otherwise — that trigger the progression to cancer⁴. “I’m sure we’ll know the answer in 20 years’ time, but we don’t know it now,” Reeves says. ■

Branwen Morgan is a freelance science writer based in Sydney, Australia.

1. Bakiri, L. & Wagner, E. F. *Mol. Oncol.* **7**, 206–223 (2013).
2. Nakagawa, H. *et al. Cancer Cell* **26**, 331–343 (2014).
3. Griffett, K. *et al. ACS Chem. Biol.* **15**, 559–567 (2013).
4. Anstee, Q. M. & Day, C. P. *Nature Rev. Gastroenterol. Hepatol.* **10**, 645–655 (2013).



to uncover tantalizing biological clues: they have identified links among hormones, inflammation and liver-cancer risk that could lead to new approaches for treating or even preventing the disease. In so doing, they may even help to clear up part of the larger mystery of why cancer, in general, afflicts more men than women.

Behaviour is partly to blame: men are more likely than women to drink, for instance, and alcohol use is a major risk factor for liver cancer. But this alone does not explain the disparity: even in rodent models of the disease, males develop more tumours than females. "It's absolutely biological. It's not lifestyle driven," says Arlin Rogers, a liver oncology researcher at the Cummings School of Veterinary Medicine at Tufts University in North Grafton, Massachusetts. "Male mice and rats don't drink more than female mice and rats."

Liver cancer almost always develops as a consequence of long-standing liver injury and inflammation caused by disorders such as hepatitis B (see page S12), hepatitis C, fatty liver disease (see page S8), alcoholism and cirrhosis. Any differences in susceptibility to inflammation between male and female livers could potentially contribute to a difference in cancer rates, says Gyongyi Szabo, a liver-disease researcher at the University of Massachusetts Medical School in Worcester — and multiple factors can amplify one another.

HORMONAL DIFFERENCES

A surprising number of fundamental differences separate male from female livers. A person's sex seems to have broad effects on how the liver develops, and hundreds of genes are expressed differently in men's versus women's livers. Many of these affect metabolism, causing male and female livers to metabolize certain drugs, toxins and carcinogens differently, says David Waxman, a cell biologist at Boston University in Massachusetts. "The liver is perhaps the best example to date of a non-reproductive tissue that shows wide differences between males and females in gene expression in general," he says. Many of these differences are caused by the pattern of growth-hormone release during development — the hormone is released in periodic pulses in males, and more steadily in females.

There are also lifelong differences in hormone levels between men and women, most obviously in levels of the sex hormones oestrogen and testosterone. The androgen receptor, which binds testosterone and other male hormones, has been implicated in liver cancer: this receptor is found in both men and women, but it occurs at higher levels in men.

The clearest evidence for a connection between the androgen receptor and liver cancer comes from male mice infected with hepatitis B virus (HBV) — the most common cause of liver cancer worldwide. In one particularly telling study, researchers genetically engineered mice to have livers that do

SEX DIFFERENCES

Luck of the chromosomes

Liver cancer strikes many more men than women — finding out why could lead to new ways of preventing the disease.

BY COURTNEY HUMPHRIES

A multitude of well-studied factors influence a person's susceptibility to cancer — genetic background, chemical exposure, diet and behaviour all contribute. But one factor that seems to play a major part in malignancy has received surprisingly short shrift from scientists: whether someone is male or female.

Some cancers, such as prostate or ovarian cancer, are by their very nature limited to men or women, and others, such as breast cancer, have a known connection to sex hormones. But even cancers with no clear association to sex disproportionately affect males or females

for reasons that no one yet fully understands. On balance, men get the worse lot: a sizeable list of cancer types are more common in men than women, as is cancer overall.

The imbalance is especially evident in hepatocellular carcinoma (HCC), the most common liver cancer (see 'Sexual inequality'). It affects twice as many men than women, and seems to be more aggressive in men, too. According to one analysis, women under 55 with HCC have a 17% greater chance of survival than men of the same age (although that difference disappears over the age of 65).

Historically, scant attention has been given to the reasons for this difference in susceptibility. But in the past decade, researchers have begun

not contain the androgen receptor¹. When they administered a chemical that is known to promote liver cancer, both the male and female mice had a lower incidence of liver tumours than did control mice. (The sex differences, although diminished, were not entirely abolished.) Chawnsang Chang, a cancer researcher at the University of Rochester Medical Center in New York and senior author of the study, notes that the androgen receptor enhances the virus's ability to replicate, and the receptor-virus interaction triggers the expression of genes that are linked to liver cancer.

Chang suggests that one way to interfere with this interaction is to therapeutically block activity of the androgen receptor in the liver. Chang founded AndroScience, based in Solana Beach, California, to develop one such drug, ASC-J9, which he believes could be useful for attacking liver cancer caused by HBV. It recently completed a phase II safety and efficacy trial as a topical acne medication but has not yet been tested as a human cancer treatment.

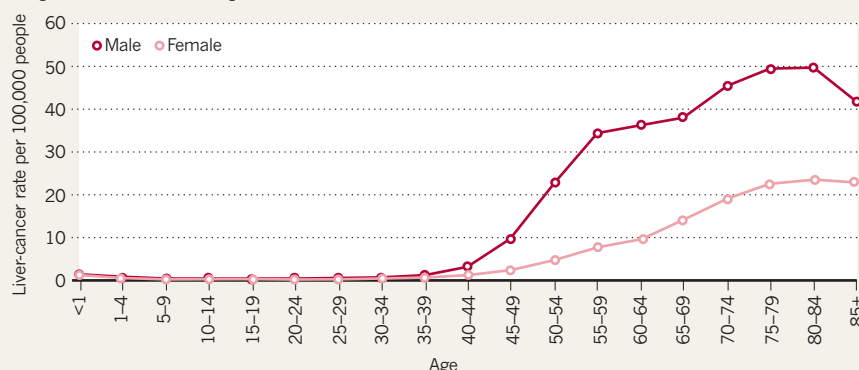
Whereas the androgen receptor seems to promote HBV-associated liver cancer, the oestrogen receptor might have exactly the opposite effect. Some studies have found that, when activated, it dampens viral replication², and oestrogen might also protect women by repressing interleukin-6, a protein that is linked to chronic liver inflammation³. Because sex hormones are involved in the early or middle stage of liver carcinogenesis, targeting them earlier in the process — for example, when someone has developed a risk factor such as an HBV infection or cirrhosis — might be a way to prevent cancer from developing in the first place, says Pei-Jer Chen, a liver-cancer researcher at National Taiwan University Hospital's Hepatitis Research Center in Taipei.

For a prevention-based hormonal approach to be feasible, Chen says, a drug must target only the liver. The 'chemical castration' hormone therapy given to men with prostate cancer comes with considerable side effects, such as hot flushes and loss of sexual function, and is too severe to give to those deemed at increased risk of liver cancer. Chen's group is currently seeking a drug that targets the liver's androgen pathway. Similarly, an oestrogen-like molecule that worked specifically in the liver could act as a preventative.

Another line of research indicates that, at least in mice, the oestrogen receptor also interacts with the proteins FOXA1 and FOXA2, which help to control various genes in the liver. Klaus Kaestner, a geneticist at the University of Pennsylvania's Perelman School of Medicine in Philadelphia, has shown that both the androgen and oestrogen receptors need FOXA proteins to bind to their molecular targets in liver cells. When mice lacking these proteins were given a carcinogen that causes liver cancer, the male mice had fewer tumours than the male controls and, conversely, the female mice had many more tumours than the female control mice⁴.

SEXUAL INEQUALITY

Data from the US National Cancer Institute show the stark difference in liver-cancer rates between men and women. The graph is based on statistics of the institute's Surveillance, Epidemiology, and End Results Program from 18 cancer registries in the United States for 2002–11.



FOXA proteins are too fundamental to the liver's basic function to serve as drug targets, but they could act as biomarkers for liver-cancer risk. Kaestner's team found that women with liver cancer often have genetic variations that stop both the FOXA proteins and the oestrogen receptor from binding to their DNA targets. Genetic variations such as these, Kaestner says, could explain why some people are more susceptible to cancer and could provide a genetic test to identify those most at risk.

THE PROLACTIN EFFECT

Although oestrogen and testosterone have garnered the most attention, another hormone also seems to be involved in liver-cancer risk. Rogers has uncovered an entirely new mechanism that might help to shield women from the disease — one that he believes could be an easy target.

Prolactin, a hormone best known for its role in stimulating milk production, is secreted by the pituitary gland and is present in all vertebrates, male and female. It has hundreds of known biological functions, ranging from regulating water balance to stimulating cell growth. Although it has been well studied in mammalian breast tissue, its role in the liver was not well known.

Rogers found that in cultured mouse liver cells and in living mice, prolactin dampens inflammation, a known precursor to HCC⁵. And mice that had been genetically engineered to lack prolactin had faster-growing tumours than did control mice when exposed to a chemical that promotes liver cancer. To determine whether prolactin had protective potential, Rogers gave male mice a drug called domperidone, which promotes prolactin production. When he later dosed them with the same cancer-promoting chemical, the domperidone-treated

mice developed fewer liver tumours.

Domperidone, which blocks the dopamine receptor, is most often used to treat nausea and vomiting; other drugs in its class are widely prescribed for schizophrenia and bipolar disorder. These medications could not be used to treat existing liver cancer, Rogers says, but they offer a way to prevent liver disease from developing into cancer. He is now working with epidemiologists to study whether people with schizophrenia who take these drugs have lower rates of liver cancer. "If we can show an association, then maybe we can sell the idea of intentionally raising prolactin," he says. "We think this might be one approach in high-risk men to reduce risk."

The prolactin study adds yet another twist to a field full of interesting findings but no consensus. "There are a number of different mechanisms that have been implicated, but at the moment no uniform picture has emerged," Waxman says. Independently, the data in each study seem clear and the resulting models solid, he says, "but how do you put it all together?"

Variations between male and female biology go far beyond the most obvious, reproductive differences. Part of the challenge, Waxman says, will be parsing those fundamental differences in tissues such as the liver and then linking them to specific disease mechanisms in males and females. This kind of increased attention to the subtler effects of one's sex could finally begin to explain why inequality prevails in not just cancer but other conditions, too — such as neurological and autoimmune diseases — even if it cannot explain inequality elsewhere. ■

Courtney Humphries is a freelance science writer based in Boston, Massachusetts.

1. Wu, M.-H. *et al. Sci. Transl. Med.* **2**, 32ra35 (2010).
2. Liu, W.-C. & Liu, Q.-Y. *World J. Gastroenterol.* **20**, 6252–6261 (2014).
3. Naugler, W. E. *et al. Science* **317**, 121–124 (2007).
4. Li, Z., Tuteja, G., Schug, J. & Kaestner, K. H. *Cell* **148**, 72–83 (2012).
5. Hartwell, H. J., Petrosky, K. Y., Fox, J. G., Horseman, N. D. & Rogers, A. B. *Proc. Natl Acad. Sci. USA* **111**, 11455–11460 (2014).



HEMIS/ALAMY

A mother and baby at a hospital in Baganuur, Mongolia. Infant vaccination against hepatitis B could reduce the country's crippling high rate of liver cancer.

VACCINES

Taking a shot at protection

A vaccine to prevent hepatitis B, the most common cause of liver cancer, has promise — and limitations.

BY KAREN WEINTRAUB

The worldwide hepatitis B virus (HBV) epidemic frequently feels personal for Indermohan Narula, a physician in Ulaanbaatar, Mongolia. Earlier this year, he saw a young woman whose liver was so scarred by HBV infection that, were she to continue her pregnancy, she would have put her life at risk. Her liver disease was untreatable, so he did what he could: he encouraged her to protect herself by ending the pregnancy and avoiding fatty foods and alcohol. “The question now is what she can do for herself and her family,” he says.

Despite such sad encounters, Narula, who is team leader of the Global Fund to Fight AIDS, Tuberculosis and Malaria in Mongolia,

sees hope for the future. Four years earlier, he helped that same woman to deliver a healthy baby boy. The infant immediately received the HBV vaccine, greatly increasing the odds that he would not develop the same disease as his mother. Stories such as this portend a healthier future for Mongolia, which currently has the world's highest rate of liver-cancer mortality — here and elsewhere, the main cause of this disease can be eliminated by the HBV vaccine.

More than 240 million people worldwide are chronically infected with HBV, according to the World Health Organization, and the virus is responsible for just over half of the world's cases of liver cancer and about 30% of all cases of cirrhosis (scarring of the liver)¹. Public-health officials are trying to protect future generations

by vaccinating as many children as possible; they have made considerable progress, but the virus is proving to be a tricky target. Not all children are getting vaccinated, and those who are still sometimes get infected. Until those gaps can be closed, and until decades have passed and immunized children grow into adulthood, HBV will remain a major threat to public health.

Early vaccination is crucial in the fight against HBV, which is generally transmitted during childbirth or through dirty needles or sex. More than 90% of adults exposed to the virus clear the infection within six months, most with few or no symptoms. But the younger that someone is exposed, the greater the odds of a chronic infection. Younger immune systems seem to be less able to fight the infection without help, says Vinh Pham, an infectious-disease specialist at the New York University Langone Medical Center. If infants are exposed to HBV during birth, immediate vaccination can erase the virus from their systems. But an unvaccinated child who picks up the virus during the first year of life will almost certainly develop a chronic infection, and 15–25% of these children will eventually die from related liver cancer or cirrhosis.

The World Health Organization has recommended universal infant vaccinations against HBV since 1992. As of 2012, 183 nations routinely vaccinate infants, and 79% of children worldwide are protected. “That was really the first vaccine that got [widespread] uptake

in the developing world,” says Paul Offit, an infectious-disease specialist at the Children’s Hospital of Philadelphia in Pennsylvania.

As with most vaccines, however, implementation rates are still below where they should be, says Neal Halsey, director of the Institute for Vaccine Safety at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. The problem is particularly acute, Halsey says, in central African countries, which have some of the highest rates of HBV infection. More vaccinations in these countries would greatly reduce rates of cirrhosis and liver cancer decades into the future.

Universal infant vaccination against HBV has proven to be hugely successful in many countries. The United States instituted such a programme in 1991; before then, about 18,000 children under the age of 10 were infected there every year. About half caught the virus from their mothers during childbirth, Offit says, and the rest were infections passed on through unprotected sex (including rape), sharing dirty needles, or casual interactions such as sharing a toothbrush or washcloth. Vaccination has now almost eliminated the infection among this age group.

INOCULATION DEFERRED

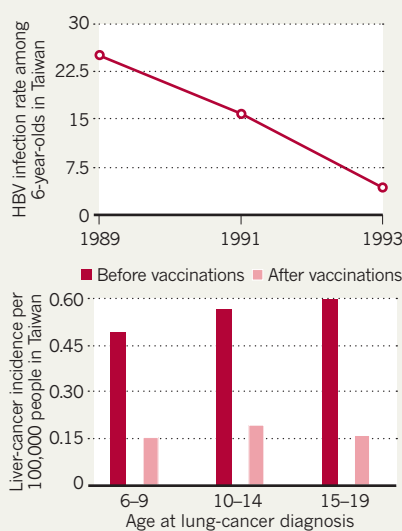
Preventive strategies are not limited to universal infant vaccination programmes — Canada and some countries in Europe vaccinate children at school, between the ages of 10 and 12. Such an approach works well in nations without a lot of mother-to-child transmission and where there is infrastructure to provide the multi-shot vaccination in adolescence. It would not, Halsey argues, be as effective in the United States, where vaccinations are not given in schools and where healthy teenagers typically see a doctor just once a year. In addition, parents in the United States have been slow to accept the idea of giving vaccines associated with sexually transmitted diseases to their children. The backlash against another cancer vaccine, one typically given to young adolescents to protect against cervical cancer caused by the human papilloma virus (HPV), has led researchers to speculate that attempts to move HBV vaccination from infancy to adolescence would elicit a similar parental reaction.

Six countries with low HBV infection rates — the United Kingdom, Denmark, Finland, Iceland, Norway and Sweden — have opted to forgo universal vaccination, choosing instead to focus on screening those at highest risk. Although there is clear consensus in the United Kingdom that childhood vaccination is important, adding an HBV shot to the schedule is complicated by politics and financial constraints, says John Edmunds, a specialist

“In the long run, the hepatitis B vaccine will be one of the most cost-effective vaccines.”

IMMUNITY BOOST

Taiwan launched a programme of universal infant vaccination against hepatitis B virus (HBV) in 1986. Since then, rates of HBV infection in the country have tumbled, along with rates of childhood liver cancer.



in infectious-disease modelling at the London School of Hygiene and Tropical Medicine. “It’s not necessarily the highest priority when compared with other changes to the vaccine programme that we could and should introduce,” he says.

In many countries, however, screening programmes trying to identify those at highest risk are simply ineffective. In the 1980s and early 1990s, the United States targeted vaccination to adults at highest risk — including sex workers, intravenous drug users and men who had sex with men — and recommended universal screening of pregnant women. But that policy failed because “people didn’t identify as being at high risk”, Halsey says, and vulnerable mothers, including sex workers, slipped through the cracks. The rates of new HBV infections in adults barely budged. It was another reminder that human behaviour can undermine even the best vaccines, and even more reason why nationwide infant immunization programmes might — at least for some countries — be the most effective option.

THE WAY FORWARD

Taiwan, which has had universal vaccination since 1986, provides perhaps the best example of the HBV vaccine’s potential benefits (see ‘Immunity boost’). Between 1977 and 1980, there were 454 cases of liver cancer among young Taiwanese people aged 5 to 29, whereas between 2001 and 2004 there were just three cases². But in adults too old to have been vaccinated, Pham says, the rate of liver cancer is essentially unchanged because people who are chronically infected rarely clear the virus.

The oldest of those vaccinated are now reaching the age of 30, and protection is largely

holding. The data suggest that booster shots are not generally needed — at least so far, says Chien-Jen Chen, a liver-cancer researcher and vice president at the Genomics Research Center at Academia Sinica in Taipei. “It seems to be that the hepatitis B vaccine has very long-term protection,” he says.

But vaccinations do not always guarantee complete immunity. A study first published in 2012 found low rates of infection — around 2% — in nearly 9,000 Taiwanese high-school students who had been vaccinated as children³. Infection rates were highest in teenagers whose mothers were also infected, suggesting that the viral load received during birth might have overwhelmed the vaccine.

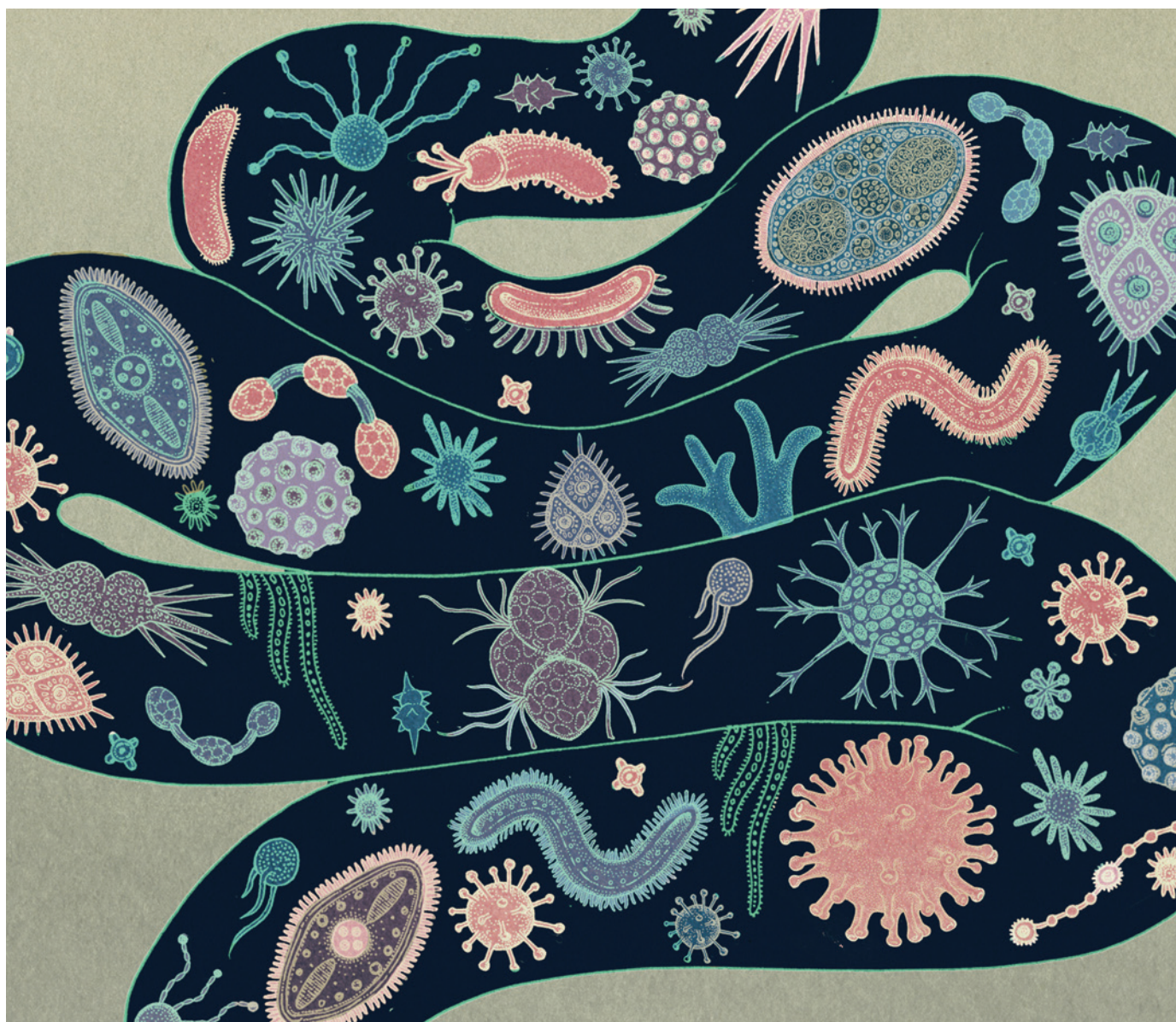
Across the Formosa Strait, in China, a universal infant immunization programme that started in 1992 is credited with protecting 80 million children from HBV infection. When vaccinations began, nearly 10% of children under the age of five had the virus. By 2006, the infection rate had dropped below 1%. But this resounding success has a potentially troubling footnote. According to a 2013 study, many of the viruses that remain in the younger population have accrued mutations — a sign that the pathogen might be evolving under the selective pressure of vaccines⁴. Doctors are watching carefully for any sign of a new strain that is broadly resistant to vaccination.

For now, researchers see no reason why even more babies around the world cannot be vaccinated. A three-dose vaccine regimen can cost US\$30 or less and rarely triggers an allergic response or any serious complication. “It’s really a very safe vaccine,” says Halsey. Despite this, some parents remain uncomfortable with vaccines, not trusting them to do more good than harm. “We begged for cancer vaccines for several decades,” Halsey says. “Parents should not be hesitant to protect their child against a relatively common cause of cancer.”

Liver cancer is an expensive disease in terms of both human lives and treatment cost (see page S4). According to Edmunds, “in the long run, the hepatitis B vaccine will be one of the most cost-effective vaccines. But it takes a very long time to achieve those gains.” Public-health systems focus their resources where they are most acutely needed. Does it make more sense to spend public money addressing a virus that will not cause liver cancer for decades or a virus such as HIV that is causing pain and suffering today? At this point, says Edmunds, that is a question each country is addressing for itself. ■

Karen Weintraub is a freelance writer based in Cambridge, Massachusetts.

1. El-Serag, H. B. *N. Engl. J. Med.* **365**, 1118–1127 (2011).
2. Chiang, C.-J., Yang, Y.-W., You, S.-L., Lai, M.-S. & Chen, C.-J. *J. Am. Med. Assoc.* **310**, 974–976 (2013).
3. Wu, T.-W., Lin, H. H. & Wang, L.-Y. *Hepatology* **57**, 37–45 (2013).
4. Bian, T. et al. *J. Virol.* **87**, 12196–12206 (2013).



KATIE SCOTT

MICROBIOME

The bacterial tightrope

Imbalances in gut bacteria have been implicated in the progression from liver disease to cancer. This insight opens the way to preventive treatments.

BY KATHERINE BOURZAC

In 2012, Eiji Hara was studying the effects of a high-fat diet on cancer risk in mice. Researchers had known for many years that obesity was associated with liver cancer; Hara wanted to find out why. After deep study of the connection, he found something unanticipated: a link to bacteria living in the gut¹.

"We never expected a connection with the microbiome," says Hara, who is a biologist at

the Japanese Foundation for Cancer Research in Tokyo. Yet when he and his colleagues exposed obese mice to a carcinogen that normally causes liver cancer and then gave them antibiotics, they found that killing the bacteria had a protective effect: the animals did not develop the disease.

The team's research, published last year, suggests that gut bacteria — which are part of the microbiome of bacteria and other microorganisms that live in and on the body — can play a crucial part in liver-cancer progression.

If further studies bear out Hara's results, they might help to answer some fundamental questions about the disease itself, and potentially lead to new ways to prevent liver cancer.

Liver cancer almost always develops in the wake of preceding problems, such as fatty liver disease (see page S8) or viral hepatitis (see page S12). But it is not clear why some people with such problems go on to develop cancer, whereas others do not — only about 3% of cases become malignant each year. Still, "liver cancer without

liver disease is very rare", says Robert Schwabe, a gastroenterologist and liver researcher at Columbia University in New York. Schwabe, Hara and others think that certain elements in the microbiome might tip the scales towards cancer. If they are right, it could be possible to stop liver disease from turning malignant by targeting microbes in the gut.

GUT FEELING

There are trillions of microorganisms in the human microbiome — they outnumber their host's cells by around ten to one — and their exact role in health and disease is only now starting to be explored. Studies have found that people with non-alcoholic fatty liver disease have a different composition of bacteria in their gut from healthy individuals^{2,3}. "We're not far along enough with this research to say why this is," says Gyongyi Szabo, who investigates the immune system and inflammation in the liver at the University of Massachusetts Medical School in Worcester.

Whatever the reason, changes in the microbiome are not sufficient to cause disease, she says. Instead, she sees an emerging picture of liver disease and cancer as a process in which various factors — including a high-fat diet, alcoholism, genetic susceptibility and the microbiome — can each contribute to the progression from minor to severe liver damage, and from severe liver damage to cancer.

It makes intuitive sense that the gut microbiome would be implicated in liver cancer, says Yu-Jui Yvonne Wan, who studies liver disease at the University of California, Davis. "Whatever comes from the gut enters the liver," she says. As blood leaves the intestines it passes through the liver, carrying nutrients extracted from our food. It also carries our own digestive chemicals as well as fragments of the bacteria that live in the digestive system (such as bits of DNA and cell walls) and their metabolic by-products. Some of these by-products help us to access energy and nutrients in food, but some are toxic. And a few might provoke an inflammatory immune response in the liver — if the inflammation becomes chronic, cancer can be the consequence.

The liver has several crucial functions, and it does more than just filter toxins and produce digestive chemicals. "The liver is an immune organ," says Richard Flavell, an immunobiologist at the Yale School of Medicine in New Haven, Connecticut.

Flavell's research suggests that the liver has an important role in immune surveillance and helps to maintain bacterial balance in the gut. Specialized cells in the liver and intestines monitor the microbiome by keeping tabs on bacterial by-products as they pass through. These cells can detect infections and help to fight them.

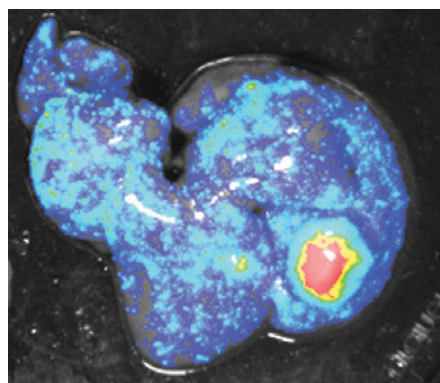
But they can also pick up on subtler changes in the bacterial populations in the gut. When certain types of bacteria become too numerous — a state called dysbiosis — the immune

system becomes activated and triggers inflammation, although at a lower level than it would for an infection. This usually goes undetected because it is not painful in the same way as arthritis or other kinds of inflammation, says Michael Karin, a molecular biologist at the University of California, San Diego, who studies the connection between inflammation and cancer. "Liver inflammation is an unappreciated, silent disease," he says.

Immune reactions aimed at fighting infectious bacteria and other foreign invaders can have off-target effects that lead to disease. Flavell's work suggests that dysbiosis causes inflammation that contributes to the worsening of non-alcoholic fatty liver disease in mice⁴. And certain cell-signalling molecules that are characteristic of inflammation can also encourage the growth of cancer cells. Now, research is emerging that suggests that dysbiosis and the immune reaction it provokes can even contribute to cancer.

Some of the earliest evidence of this comes from Schwabe, who studies the connection between liver injury and cancer. His work indicates that the immune response to the microbiome drives the progression of cancer in mice. "We discovered that cells in the liver express [immune] receptors that bind to bacterial products," Schwabe says, and the main culprits are lipopolysaccharides, large molecules that are found in the cell walls of many bacteria. When his group genetically engineered mice to lack Toll-like receptor 4, an immune receptor that lipopolysaccharides often bind to, the animals did not develop liver cancer.

Like Hara, Schwabe then went on to connect cancer progression to the microbiome by testing two groups of mice: one in which the animals had been isolated from birth so that they were completely bacteria-free, and another in which they had been dosed with strong antibiotics. All of the animals were treated with the same liver-damaging carcinogen as in the first experiment, but both groups remained cancer-free. For the disease to occur, the mice needed all three factors: an active immune pathway, the harmful bacteria and the carcinogen⁵.



Following provocation from gut-bacteria metabolites, an obese mouse liver expresses cancer-associated proteins (red fluorescence).

Targeting these newly discovered pathways therapeutically is a difficult proposition, however. Knocking out Toll-like receptors, which are a vital part of the immune system, is not practical. Nor is a broad attack on all bacteria that make lipopolysaccharides, because both harmful and helpful bacteria (such as *Lactobacillus* strains found in yogurt) have these molecules in their cell walls.

Because the most harmful strains of bacteria have not yet been pinpointed, and because the specific details of the inflammatory process are still unclear, researchers need more information before they can target this process therapeutically. "Right now we're working on finding

It makes intuitive sense that the gut microbiome would be implicated in liver cancer.

the exact mechanism of cancer progression and the exact cell type in the liver that is targeted by lipopolysaccharides," Schwabe says.

He thinks that at least part of this mechanism involves

disruption in the balance of the various species of bacteria in the gut. An out-of-balance microbiome promotes a constant state of inflammation, which can contribute to cancer progression, Schwabe says. This aligns with the picture that is emerging of cancer, in general, as an inflammatory process: the same immune reactions that help the body to fight infection and disease can also promote unchecked cell growth.

OUT OF BALANCE

Another common bacterial product might also serve as a connection between the microbiome and liver cancer. The liver generates bile acids, which are stored in the gall bladder and are released into the intestines to digest fatty foods. They act as a kind of digestive detergent, helping the body to take up dietary fat along with fat-soluble vitamins. The acids then get absorbed by the intestines and passed back to the liver, which sends them back to the gall bladder.

Some gut bacteria chemically alter these bile acids to produce secondary bile acids, which are toxic to humans and even to certain bacteria. The most common of these secondary bile acids is deoxycholic acid (DCA). A certain amount of DCA is present in everyone's digestive system, and at low levels it does not seem to have ill effects. It carries fat in the same way as our own bile acids and follows the same path from gut to liver to gall bladder. But it can cause DNA damage, which in turn can lead to cancer.

An abundance of DCA can also lead to dysbiosis, which not only is a link in the cancer-progression chain, but also has been connected to obesity. Some of the earliest research on the human microbiome, published in 2006, demonstrated that the balance of gut bacteria in obese people is different from that in people of healthy weight. In particular, obese people tend

to have greater numbers of the bacteria that produce DCA and other secondary bile acids. The result is a positive-feedback loop, says Wei Jia, a pharmaceutical chemist at the University of Hawaii in Honolulu. DCA kills certain bacteria, which leads to a greater preponderance of the DCA-producing strains. “It really changes the composition of the microbiota in the intestine and can cause dysbiosis,” Jia says.

Hara’s work in Tokyo, which connected liver cancer to the microbiome of mice on a high-fat diet, also points to DCA¹. Hara originally intended to study the connection between obesity and liver cancer, so he treated his mice with a carcinogen at birth and fed half of them a high-fat diet. None of the animals on a normal diet developed liver cancer, whereas all of those on a high-fat diet developed the disease. When Hara treated the overweight mice to decrease DCA production, the cancer incidence fell, too.

This line of research points to the microbiome as one potential link between obesity and liver-cancer risk (see ‘Bacterial links’). And, much like Schwabe’s work, Hara’s results indicate that several factors converge to promote cancer: in this case, bacteria, diet and carcinogen exposure. Here, too, the ability to stave off the disease seems to depend on maintaining the appropriate microbial balance. Overweight mice and people have a different composition of gut microbiota from their lighter counterparts, and they have higher levels of DCA, too.

THE ROAD TO PREVENTION

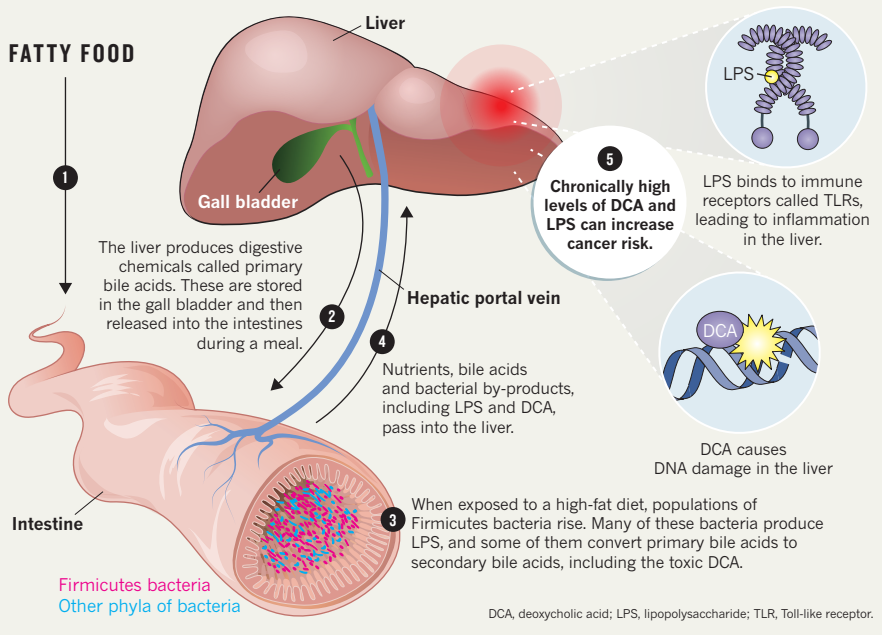
The complexity of the human microbiome presents considerable challenges for researchers. Improvements in genetic technologies have spurred progress in this area, and the US National Institutes of Health has invested US\$200 million so far in the Human Microbiome Project, a large-scale genomics and analysis project that aims to characterize our microbial inhabitants. However, most work on the microbiome focuses on entire ecosystems of bacteria, and genomic sequencing is done on entire populations, making it difficult to identify individual species.

Hara hopes to identify specific bacteria that might be implicated in cancer progression, but it is a difficult proposition. Most of the bacteria living in the gut are anaerobic and cannot be grown using conventional laboratory methods. “We’re trying hundreds of culture conditions but we can’t isolate the bacteria yet,” he says.

To prove that the microorganism–cancer link exists, and to develop preventive therapies, researchers must first demonstrate that the microbial effects seen in mouse studies also occur in humans. Hara’s group is now collecting clinical samples of blood and stool from 1,000 healthy volunteers, 1,000 people with colon cancer and 50 people who developed liver cancer after having non-alcoholic fatty liver disease. He plans to analyse bacterial genes in stool and DCA concentration in the blood, and expects results in about two years.

BACTERIAL LINKS

The gut and liver are intimately connected. Bacterial populations living in the gut change their composition in response to diet, and such bacterial activity might contribute to liver-cancer risk and progression.



If Hara and his colleagues can identify the bacteria that contribute to liver cancer, tests could be developed to identify people whose microbiomes put them at risk. “Then we can screen people using PCR or antibodies,” he says.

However, not everyone is convinced that individual bacterial species are to blame. Some researchers point out that dysbiosis, and therefore cancer risk, involves multiple strains of bacteria. And the bacterial mix can vary from person to person, meaning it is unlikely that scientists can pin all responsibility on a single species. They instead advocate targeting specific genes, rather than specific bugs. Phillip Hylemon, a microbiologist and immunologist at Virginia Commonwealth University in Richmond, is working on ways to block bacterial secondary-bile-acid synthesis pathways. “We’ve isolated some of the enzymes involved, and we have three-dimensional structures of some of them,” he says. Pharmaceutical chemists can use these structures to develop molecules that are shaped specifically to bind to and inactivate these enzymes, thereby preventing secondary-bile-acid formation and reducing liver-cancer risk.

Others are looking for ways to promote the growth of healthy bacterial strains rather than target the bad ones. Jia, who is analysing the composition of people’s bile acids, thinks it might be possible to get microbiomes back in balance by treating them with ‘good’ bacteria

that have been genetically engineered to produce an enzyme that converts secondary bile acids back to their primary form.

There is also some early clinical evidence that specially formulated probiotics — cocktails of good bacteria — can bump the microbiome back into balance. Hylemon and his colleagues gave people with cirrhosis a probiotic containing *Lactobacillus* bacteria and found that their blood markers of inflammation decreased along with their cognitive dysfunction (a common symptom of cirrhosis)⁶. Although the study was not designed to evaluate cancer risk, it does show that delivering bacteria to the gut can have positive therapeutic effects on the liver.

It is still early days for this research. “One-third of the metabolites in the blood are coming from gut bacteria, and we’ve learned this only in the past few years,” Hylemon says. But although the field is still in its infancy, it is already providing researchers with a whole new way to think about cancer. If physicians can identify people who are at the greatest risk of developing liver cancer — those with underlying liver damage and a genetic predisposition to disease — they might be able to harness the microbiome and stop the cancer before it starts. ■

Katherine Bourzac is a freelance science writer in San Francisco, California.

1. Yoshimoto, S. *et al.* *Nature* **499**, 97–101 (2013).
2. Mouzaki, M. *et al.* *Hepatology* **58**, 120–127 (2013).
3. Raman, M. *et al.* *Clin. Gastroenterol. Hepatol.* **11**, 868–875 (2013).
4. Henao-Mejia, J. *et al.* *Nature* **482**, 179–185 (2012).
5. Dapito, D. H. *et al.* *Cancer Cell* **21**, 504–516 (2012).
6. Bajaj, J. S. *et al.* *Aliment. Pharmacol. Ther.* **39**, 1113–1125 (2014).

PERSPECTIVE



Incision revision

Antiviral medication is leading surgeons to re-evaluate how they treat hepatitis-induced liver cancer, writes **Myron Schwartz**.

Hepatocellular carcinoma (HCC), a form of cancer that begins in the liver, is the second-most-common cause of cancer-related death around the world. HCC is nearly always the result of an underlying chronic liver disease, usually viral hepatitis, and the choice and outcome of treatment depend on both the extent of the tumour and the overall condition of the liver. When cancer is confined to just part of the liver and the function of the organ remains normal, removing the diseased area (a procedure called resection) is the preferred treatment and can be curative. But if the organ has been damaged too badly by underlying disease, the only option might be to remove the entire liver and replace it with a donated one — a liver transplant. This procedure can be life saving, but not only does it carry a high risk for the patient, there are also too few donated organs to meet the demand.

However, thanks to recent advances in the treatments for viral hepatitis, HCC is not always an immediate threat. These treatments, combined with the development of better ways to diagnose and treat liver tumours at early stages, are changing the way that surgeons approach liver cancer: an increasing number of people can avoid liver transplants, and some with small tumours can be cured without any surgery at all.

Worldwide, the most common disease underlying HCC is infection with hepatitis B virus (HBV). Since the introduction of effective antiviral treatment for HBV in 1998, the survival rates of people with HBV-related liver cancer have dramatically improved. Treating the underlying HBV infection not only stabilizes or even improves liver function, thereby enabling people to better tolerate surgery, it also reduces the incidence of new HCC after successful resection by 70% or more.

In Western countries, however, the predominant disease underlying HCC is hepatitis C virus (HCV). In contrast to HBV, HCV has proven difficult to treat, particularly in people with HCC who have extensively scarred livers because the available HCV treatments have been both toxic and only marginally effective. At the Mount Sinai Medical Center in New York City, the five-year survival rate after both resection and transplantation for HCV-related HCC has been 10–15% worse than for HBV-related HCC, primarily owing to cancer recurrence after resection and progressive damage to the new liver as a result of HCV recurrence.

HOPE FOR HEPATITIS C

The good news is that with the approval last year of a new antiviral medicine called sofosbuvir, HCV infection has suddenly become easy to cure. This landmark development has important implications not just for HCV but also for treating and even preventing liver cancer.

HCC emerges only late in the course of HCV infection, after years of inflammation and constant regeneration have led to a substantial build-up of scar tissue (fibrosis) and a constellation of molecular changes that culminate in cancer. The old, toxic HCV treatment rarely cured people, but when it did, it reduced the risk of developing HCC by a factor of 5 to 6; sofosbuvir, if started before the development of

cirrhosis, could potentially reduce that risk to almost zero. However, curing HCV does not eliminate other risk factors for HCC, such as alcohol abuse, obesity and diabetes, all of which are common among people with HCV in the United States. These factors, together with male gender, older age, and the extent of fibrosis, can contribute to ongoing risk of HCC after HCV clearances.

The advent of effective HCV treatment will almost certainly improve the results of surgery in people with HCV-related HCC. The success rates for liver transplants are highly likely to reach the level seen in people on antiviral therapy with HBV-related HCC, who have a five-year survival rate of 75–80%. For resection, predicting the degree of improvement is more difficult. Both recurrence of the original HCC and development of new HCC within the liver have been more common after resection in people with HCV than in those with HBV, and eliminating the virus can only reduce recurrence. In people with HCV-associated

HCC who are good candidates for either transplant or resection, there is little difference in success rates between the two approaches. In fact, when survival is calculated from the time that people are put on the transplant waiting list instead of from the time the transplant is performed, survival after resection is often superior to that for transplant: people can die or become ineligible for transplant during the often-prolonged waiting time for a matched donor to become available.

Donor livers are a scarce resource. When deciding how to best use them, we must consider the benefit of a transplant relative to the alternatives. For people with HCC who have borderline liver function, a strategy of initial non-surgical HCC treatment to forestall tumour progression and a 12-week course of antiviral therapy might improve the condition of their livers and thereby

render them more suitable for resection. Historically, 76% of people with HBV-related HCC treated surgically at our centre have undergone resection compared with only 40% with HCV-related HCC; with use of sofosbuvir, the latter figure is likely to rise and results should improve.

As more people become eligible for potentially curative liver-cancer treatments, it is now a public-health imperative to promote screening and surveillance programmes aimed at identifying HCC at an early stage, when it is amenable to curative therapy.

Beyond helping patients who already have HCC, however, the most important effect of the new HCV drugs will be preventing liver cancer from occurring in the first place. Clearing HCV infection before people develop significant fibrosis effectively eliminates their risk of HCV-related HCC. The challenge now will be identifying the people who are already infected; in the United States, a recent government mandate for HCV screening is an important step in the right direction. We must mount a campaign to raise public awareness about this potentially devastating illness, which can now be simply diagnosed and cured. ■

Myron Schwartz is director of liver surgery at Mount Sinai Medical Center's Recanati/Miller Transplantation Institute in New York City.

IT IS NOW A
PUBLIC-HEALTH
IMPERATIVE
TO PROMOTE
SCREENING
AND
SURVEILLANCE
PROGRAMMES.